



**The Effects of Stereotype Threat on
Foreign Language Performance
through the Mediating Roles of Self-Efficacy and Anxiety**

Gulsah Kutuk

Faculty of Education

Edge Hill University

This thesis is submitted to the Department of Education, Edge Hill University, in partial
fulfilment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

June 2019

I. Declaration

I hereby declare that this thesis is entirely my own work carried out under the normal terms of supervision. I confirm that this work has not been submitted for any other degree or diploma at Edge Hill University or any other institution. The work gained ethical approval from the Faculty of Education at Edge Hill University and has been conducted in accordance with the faculty's Ethical Guidance. The sources drawn upon have been fully acknowledged throughout the thesis.

Signature: 

II. Acknowledgements

First and foremost, I would like to express my deepest appreciation and gratitude to my dear supervisors, Professor Dave Putwain, Dr Linda Kaye and Dr Bethan Garrett for their continuous support throughout the duration of my doctoral study. Without their guidance, motivation and inspiration, I would not be able to complete this thesis.

Also, I would like to take this opportunity to thank all the participants who volunteered to be part of this work. I felt blessed with their readiness to help and support me whenever I needed them. They definitely played a crucial role in making this research project possible.

Many thanks to all my lovely friends who have always offered their social support to me for years now. Thanks to them, although I am the only child in my family, I have never felt what it is like to have no brothers or sisters. I believe my life would never be such meaningful without their friendship which is a special gift for me.

I owe thanks to my fiancé, Yusuf as well. Undertaking a PhD can be a lonely endeavour, but Yusuf was always there for me giving me the strength to continue my work and making my life easier in every sense throughout my PhD journey. Therefore, I would like to thank him for not letting me feel alone and lonely during this challenging experience.

And finally, my special thanks go to my lovely parents, Nuray and Huseyin who have been a constant source of emotional support to me throughout this tough experience. I could not have imagined a life without their unwavering love, support and encouragement. Therefore, I would like to dedicate this thesis to my mother and father who mean all the world to me.

III. Abstract

Background: Gender stereotyping of academic domains has long been a subject of debate in the field of education. Researchers believe that gender stereotyping of academic domains is an issue because how an academic subject is perceived by females and males can influence their achievement-related perceptions including their self-efficacy and anxiety. In alignment with this argument, several studies have found that males and females tend to favour and be more confident in the academic subjects which are believed to be more appropriate for their gender.

Aims: Although vital to academic achievement, a substantial body of research focusing on the impact of gender stereotyping of academic subjects is mainly concerned with females and their underachievement in certain subjects such as maths and science. Conversely, there is little attention to males and their performance in academic fields which are mostly associated with females. This thesis, therefore, aimed to explore the concept of gender stereotyping in respect of males and their performance in foreign language learning which, in some language learning environments, is believed to be a female domain. The research investigated the extent to which any existing gender stereotypes were linked to foreign language learners' performance via the mediating roles of language self-efficacy and anxiety.

Sample: Overall, 1140 Turkish adult learners (509 females, 631 males) studying English as a foreign language at university level and 17 English as a foreign language teachers (7 males, 10 females) were recruited across three studies as well as a preceding pilot study.

Method: A mixed methods approach incorporating self-report questionnaires, interviews and experimental methods was adopted in this research. Study 1 employed a questionnaire design which examined whether there was a link between language learners' gender stereotyped beliefs about foreign language learning and their self-efficacy, anxiety, and performance. Study 2 took an interview approach with language teachers and learners and explored the extent to which language teachers, as an agent of socialisation, played a role in sustaining or legitimising any existing gender stereotyped beliefs. Finally, Study 3 experimentally investigated the impact of stereotype

threat pertaining to learning another language upon male language learners' performance via their self-efficacy and anxiety.

Results: The results in Study 1 and 2 confirmed that Turkish learners of English as a foreign language held the belief that females were better at foreign language learning in a number of ways (e.g., hard work, perseverance) compared to males. Study 2 also revealed that these beliefs were formed prior to university and by families and the society rather than foreign language teachers. Although prevalent, the belief that females were better language learners was not found to be linked to males' performance through their self-efficacy and anxiety in Study 1. Study 3 demonstrated that male learners in experimental conditions performed worse in English listening test than the control group suggesting that the phenomenon of stereotype threat might be true for males in foreign language learning environments. However, consistent with Study 1, the effects of gender stereotypes were not found to be transmitted through reception self-efficacy or anxiety.

Conclusion: The empirical research presented in this thesis is among few studies examining the phenomenon of gender stereotyping of academic subjects in terms of males and their foreign language performance. The outcomes of this study contribute to the field of second language studies as it enabled a deeper understanding of the impact of gender stereotyped beliefs on language learners' self-efficacy, anxiety and performance.

Keywords: Foreign language learning, gender stereotypes, self-efficacy, anxiety, stereotype threat, language performance

IV. Table of Contents

I. Declaration	ii
II. Acknowledgements	iii
III. Abstract	iv
V. List of Figures	xi
VI. List of Tables	xiii
VII. List of Abbreviations	xv
VIII. Glossary of Terms	xvii
CHAPTER 1: INTRODUCTION	1
1.0. Chapter Outline	1
1.1. Gender stereotyping of Academic Domains	1
1.2. Gender Roles and Expectations in Turkey	3
1.3. The Current Thesis	5
1.4. Chapter Summary	7
CHAPTER 2: LITERATURE REVIEW	8
2.0. Chapter Outline	8
2.1. Common Terminologies in Second Language Acquisition	8
2.1.1. Second Language Acquisition vs. Second Language Studies	8
2.1.2. Second Language Acquisition vs. Second Language Learning	9
2.1.3. Second Language Acquisition vs. Foreign Language Learning	10
2.2. The Four Language Skills	12
2.2.1. Listening	13
2.2.2. Reading	15
2.2.3. Speaking	16
2.2.4. Writing	18
2.3. The role of Individual Differences in L2 learning	20
2.3.1. Gender, Gender Stereotypes and L2 learning	20
2.3.1.2. Agents of Gender Socialisation	26
2.3.2. L2 Self-Efficacy	36
2.3.3. L2 Anxiety	43
2.4. Devising the Initial Research Questions	50
2.5. Chapter Summary	53
CHAPTER 3: METHODOLOGY	54
3.0. Chapter Outline	54
3.1. Critical Realism as a Research Paradigm	54

3.1.1.	Ontological and Epistemological Assumptions	60
3.2.	Mixed Methods Research and Data Collection Methods	63
3.2.1.	Definition of Mixed Methods Research	63
3.2.2.	Rationale for Mixed Methods Research	65
3.2.3.	Mixed Methods Research Design	66
3.3.	Setting and Participants	69
3.3.1.	The Context of Turkey	69
3.3.2.	Sampling Strategy	72
3.3.3.	Sample Size	75
3.4.	Data Collection	77
3.4.1.	Study 1: Quantitative Approach (Closed-Ended Questionnaires)	77
3.4.2.	Study 2: Qualitative Approach (Semi-Structured Interviews)	79
3.4.3.	Study 3: Quantitative Approach (Experiments)	82
3.5.	Data Analysis	85
3.5.1.	Close-Ended Questionnaires	85
3.5.2.	Semi-Structured Interviews	86
3.5.3.	Experiments	87
3.6.	Data Integration	87
3.7.	Ethical Considerations	88
3.7.1.	Voluntary Informed Consent	88
3.7.2.	Right to Withdraw	89
3.7.3.	Anonymity, Confidentiality and Privacy	89
3.7.4.	Openness and disclosure	89
3.8.	Chapter Summary	90
CHAPTER 4:	STUDY 1	92
4.0.	Chapter Outline	92
4.1.	Introduction	92
4.2.	Research Aims and Questions of Study 1	95
4.3.	Pilot Study	95
4.3.1.	Back-Translation	96
4.3.2.	Participants	97
4.3.3.	Data collection Instruments and Procedures	97
4.3.4.	Results	103
4.4.	Main Study	115
4.4.1.	Participants	115
4.4.2.	Data Collection Instruments	116

4.4.3. Results	116
4.5. Structural Equation Modelling (SEM) Results	137
4.5.1. The Relationship between Learners' Gender Stereotyped Beliefs about Language Learning and their Language Performance	138
4.5.2. The Relationship between Learners' Perceptions about their Language Teachers' Gender Stereotyped Beliefs and Language Performance 144	144
4.6. Discussion	149
4.6.1. Limitations and Suggestions for Future Research	152
4.6.2. Original Contribution to Knowledge	153
4.7. Chapter Summary	154
CHAPTER 5: STUDY 2	156
5.0. Chapter Outline	156
5.1. Introduction	156
5.2. Research Aims and Questions of Study 2	159
5.3. Method	159
5.3.1. Participants	159
5.3.2. Data Collection Procedure	160
5.3.3. Data Analysis	166
5.4. Results	167
5.4.1. Theme 1: Dispositional Factors	168
5.4.2. Theme 2: Individual Factors	173
5.5. Discussion	176
5.5.1. Dispositional Factors	176
5.5.2. Individual Factors	179
5.5.3. Limitations and Suggestions for Future Research	180
5.5.4. Original Contribution to Knowledge	181
5.6. Chapter Summary	181
CHAPTER 6: STUDY 3	183
6.0. Chapter Outline	183
6.1. Introduction	183
6.2. Research Aims and Questions of Study 3	186
6.3. Method	187
6.3.1. Participants	187
6.3.2. Stereotype Threat Manipulations	187
6.3.3. Data Collection Instruments	188
6.4. Procedure	190

6.5. Results	191
6.5.1. Descriptive Results.....	191
6.5.2. Gender Stereotype Endorsement	192
6.5.3. Stereotype Manipulation Check – L2 Learning	195
6.5.5. L2 Reception Self-Efficacy and L2 Listening Anxiety	197
6.5.6. The Mediation Analysis	198
6.6. Discussion	202
CHAPTER 7: GENERAL DISCUSSION	207
7.0. Chapter Outline.....	207
7.1. Summary and Integration of Findings.....	207
7.1.1. Gender Socialisation and Gender Stereotyping of L2 learning	210
7.1.2. The Mediating Role of L2 Anxiety and L2 Self-efficacy	213
7.2. Thesis Limitations and Suggestions for Future Research	215
7.2.1. The Sample	215
7.2.2. Data Collection Instruments.....	216
7.3. Contribution to Knowledge	217
7.3.1. Gender Stereotyping of Academic Domains	217
7.3.2. The MLCAS	218
7.3.3. The QSLL.....	219
7.4. Thesis Conclusions.....	220
APPENDICES.....	252
Appendix 3.1. Ethical Approval Letter	252
Appendix 3.2. Participant Information Sheets.....	253
Appendix 3.3. Participant Consent Forms	257
Appendix 3.4: Participant Debrief Forms	258
Appendix 4.1. Pilot Study Materials	259
Appendix 4.1.1. Part 1. Background Questionnaire.....	259
Appendix 4.1.2. Part 2: Learners' Gender Stereotyped <i>Perceptions of Subjects</i>	260
Appendix 4.2. Multidimensional Language Class Anxiety Scale (MCLAS)	262
Appendix 4.3. Common Reference Levels: self-assessment grid	263
Appendix 4.4. Questionnaire of Self-Efficacy Beliefs in Learning a New Language (QSLL)	267
Appendix 4.5. Revised Version of Learners' Perceptions About the Academic Subjects.....	268
Appendix 4.6. Revised Version of <i>Questionnaire of Language Learners' Gender Stereotyped Beliefs</i>	269

Appendix 4.7. Revised Version of Questionnaire of Self-Efficacy Beliefs in Learning a New Language (QSLL).....	271
Appendix 5.1. Interview Questions for L2 Teachers	271
Appendix 5.2. Interview Questions for L2 Learners	273
Appendix 6.1.: Preliminary English Listening Test.....	276

V. List of Figures

Chapter 2

Figure 2.1. <i>Self-as-target Threats</i>	33
Figure 2.2. <i>Group-as-target Threats</i>	35
Figure 2.3. <i>Conceptual overview of the programme of research</i>	51
Figure 2.4. <i>Self-concept threat vs. Group-concept threat</i>	52

Chapter 3

Figure 3.1. <i>Four Versions of Paradigms</i>	55
Figure 3.2. <i>Four Levels for Developing a Research Study</i>	57
Figure 3.3. <i>Causality in Critical Realism - Adapted from Blaikie and Priest (2017)</i>	61
Figure 3.4. <i>The Convergent Parallel Design of the Current Study</i>	68
Figure 3.5. <i>Six Proficiency Levels in the CEFR</i>	71
Figure 3.6. <i>Description of Seven Geographical Regions of Turkey</i>	73
Figure 3.7. <i>Sampling Stages in the Current Study</i>	74
Figure 3.8. <i>Percentages of Internal Migration to Istanbul</i>	74

Chapter 4

Figure 4.1. <i>Two-factor model of QLGB</i>	120
Figure 4.2. <i>Lower Order Correlated Residuals. For simplicity, the relations between corresponding residuals were omitted</i>	125
Figure 4.3. <i>Higher Order Correlated Residuals. For simplicity, the relations between corresponding residuals were omitted</i>	125
Figure 4.4. <i>Lower Order L2 Anxiety with Three Method Factors.</i>	125
Figure 4.5. <i>Bifactor model with correlated residual variance. For simplicity, the relations between corresponding residuals were omitted.</i>	126
Figure 4.6. <i>Two-factor model for QSLL</i>	135
Figure 4.7. <i>The SEM Model</i>	138
Figure 4.8. <i>The moderation model for learners' gender stereotyped beliefs</i>	139
Figure 4.9. <i>The SEM Model for Learners' Gender Stereotyped Beliefs</i>	142
Figure 4.10. <i>Standardised path coefficients of the hypothesised model for males</i>	143
Figure 4.11. <i>The Moderation Model</i>	144

Figure 4.12. <i>The SEM Model for Students' perceptions about their Language Teachers' Gender Stereotyped Beliefs</i>	147
Figure 4.13. <i>Standardised path coefficients of the hypothesised model for males</i>	148

Chapter 5

Figure 5.1. <i>Themes and sub-themes emerged in the thematic data analysis</i>	168
--	-----

Chapter 6

Figure 6.1. <i>One-way ANOVA results for the English Listening Test Scores across the groups</i>	196
Figure 6.2. <i>The Mediational Model</i>	199
Figure 6.3. <i>The Mediational Model - The difference between the self-as-target and group-as-target conditions</i>	199

Chapter 7

Figure 7.1. <i>The Hypothesis of The Current Research</i>	208
Figure 7.2. <i>The Convergent Parallel Design of the Current Research</i>	209

VI. List of Tables

Chapter 2

<i>Table 2.1. Summary of the Common SLS Concepts.....</i>	<i>11</i>
---	-----------

<i>Table 2.2. Summary of stereotype threat literature examining academic performance.....</i>	<i>29</i>
---	-----------

Chapter 3

<i>Table 3.1. The Three Ontological Domains of Critical Realism</i>	<i>60</i>
---	-----------

<i>Table 3.2. Common Reference Levels: Global Scale.....</i>	<i>72</i>
--	-----------

Chapter 5

<i>Table 4.1. The original class-related items created by Pekrun et al. (2005)</i>	<i>100</i>
--	------------

<i>Table 4.2. The Structure of the MLCAS</i>	<i>101</i>
--	------------

<i>Table 4.3. Frequency analysis of demographics and foreign language background of the participants</i>	<i>103</i>
--	------------

<i>Table 4.4. Descriptive Statistics of Learners' gender Stereotyped Perceptions of Subjects - The Pilot Study.....</i>	<i>105</i>
---	------------

<i>Table 4.5. Pattern Matrix for the QSL.....</i>	<i>107</i>
---	------------

<i>Table 4.6. QLGB Statistics – The Pilot Study.....</i>	<i>108</i>
--	------------

<i>Table 4.7. Descriptive Statistics of the individual items in QLGB.....</i>	<i>109</i>
---	------------

<i>Table 4.8. Exploratory Factor Analysis Results – The Pilot Study Data.....</i>	<i>110</i>
---	------------

<i>Table 4.9. Descriptive results of the domains of the MLCAS.....</i>	<i>112</i>
--	------------

<i>Table 4.10. Pattern Matrix for the QSL.....</i>	<i>114</i>
--	------------

<i>Table 4.11. The QSL Statistics – The Pilot Study.....</i>	<i>115</i>
--	------------

<i>Table 4.12. Frequency analysis of demographics and foreign language background of the participants - The Main Study</i>	<i>117</i>
--	------------

<i>Table 4.13. Descriptive Statistics of Learners' gender Stereotyped Perceptions of Subjects.....</i>	<i>118</i>
--	------------

<i>Table 4.14. Standardised Loadings for the Two-factor Solution – The Main Study....</i>	<i>121</i>
---	------------

<i>Table 4.15. Goodness of fit indices of the QLGB - the Main Study.....</i>	<i>122</i>
--	------------

<i>Table 4.16. Scale Statistics of the QLGB – The Main Study</i>	<i>122</i>
--	------------

<i>Table 4.17. Standardised factor loadings for the subscales of the MLCAS – The Main Study</i>	<i>127</i>
---	------------

<i>Table 4.18. The goodness of fit indices for the variables in the MLCAS – The Main Study</i>	<i>129</i>
--	------------

<i>Table 4.19.</i> The goodness of fit indices for all hypothesised models for the MLCAS - Main Study.....	129
<i>Table 4.20.</i> Standardised factor loadings for the bifactor model with correlated residual variance – The Main Study	130
<i>Table 4.21.</i> Item and Scale Statistics for Model 4 – The Main Study.....	132
<i>Table 4.22.</i> The domains of the MLCAS - Correlation analysis	133
<i>Table 4.23.</i> Goodness of fit indices of the QSLI - the Main Study	134
<i>Table 4.24.</i> Standardised Loadings for the Two-factor Solution – The Main Study....	134
<i>Table 4.25.</i> Scale Statistics of the QSLI – The Main Study	137
<i>Table 4.26.</i> The Chi-square difference test between the baseline model and the constrained model	140
<i>Table 4.27.</i> The Chi-square difference test between the baseline model and the model for L2 anxiety	140
<i>Table 4.28.</i> The Chi-square difference test between the baseline model and the model for L2 Self-efficacy	140
<i>Table 4.29.</i> The Chi-square difference test between the baseline model and the constrained model	145
<i>Table 4.30.</i> The Chi-square difference test between the baseline model and the model for L2 anxiety	146
<i>Table 4.31.</i> The Chi-square difference test between the baseline model and the model for L2 Self-efficacy	146

Chapter 6

<i>Table 6.1.</i> Frequency analysis of demographics and foreign language background of the participants - Experimental Study.....	192
<i>Table 6.2.</i> Descriptive Statistics of Participants' Gender Stereotyped Perceptions of Academic Subjects	193
<i>Table 6.3.</i> Univariate Test Results – Gender Stereotype Endorsement	194
<i>Table 6.4.</i> Univariate Test Results – Stereotype Manipulation Check	196
<i>Table 6.5.</i> Descriptive statistics for L2 self-efficacy, L2 listening anxiety and L2 performance	198
<i>Table 6.6.</i> Direct, Indirect and Total Effects and 95% Confidence Intervals for the Mediation Model – Self-as-target vs. Group-as-target.....	200
<i>Table 6.7.</i> Direct, Indirect and Total Effects and 95% Confidence Intervals for the Mediation Model – Self-as-target vs. Control group	201

VII. List of Abbreviations

ABIC:	Akaike's Bayesian Information Criterion
AEQ:	The Achievement Emotions Questionnaire
AIC:	The Akaike information criterion
ANOVA:	Analysis of variance
BERA:	British Education Research Association
CEFR:	The Common European Framework of Reference for Languages
CFA:	Confirmatory Factor Analysis
CFI:	Comparative Fit Index
CI:	Confidence Interval
DF:	Degrees of Freedom
DV:	Dependent Variable
EFA:	Exploratory Factor Analysis
EFL:	English as a Foreign Language
ELT:	English Language Teaching
EMI:	English as a Medium of Instruction
FIML:	Full Information Maximum Likelihood
FLCAS:	Foreign Language Classroom Anxiety Scale
FLL:	Foreign Language Learning
FRE:	The Framework for Research Ethics
HE:	Higher Education
ID:	Individual Difference
IFI:	Incremental Fit Index
IV:	Independent Variable
L1:	Mother Tongue/Native Language
L2:	Foreign/Second/Additional language
LA:	Listening Anxiety
LCA:	Language Class Anxiety
LLCI:	Lower Level Confidence Interval
MLCAS:	The Multidimensional Language Anxiety Scale
MMR:	Mixed Methods Research
MTF:	Multi-Threat Framework
QESE:	Questionnaire of English Self-Efficacy
QLGB:	Questionnaire of Language Learners' Gender Stereotyped Beliefs
QSL:	Questionnaire of Self-Efficacy Beliefs in Learning a New Language
RA:	Reading Anxiety

RMSEA:	Root Mean Square Error of Approximation
SA:	Speaking Anxiety
SEM:	Structural Equational Modelling
SIT:	Social Identity Theory
SLA:	Second Language Acquisition
SLL:	Second Language Learning
SLS:	Second Language Studies
SLWAI:	Second Language Writing Anxiety Inventory
SPSS:	Software Package for Statistical Analysis
SRMR:	Standardised Root Mean Square Residual
ST:	Stereotype Threat
STEM:	Science, Technology, Engineering and Maths
T-EMI:	Turkish and English as a Medium of Instruction
TLI:	Tucker–Lewis Index
TMI:	Turkish as a Medium of Instruction
TURKSTAT:	Turkish Statistical Institute
ULCI:	Upper Level Confidence interval
WA:	Writing Anxiety
WAT:	Writing Apprehension Test

VIII. Glossary of Terms

Critical Realism:	A philosophical approach associated with Roy Bhaskar (1944–2014). Critical realism adopts epistemological relativism and judgemental rationality. It acknowledges that knowledge cannot be universal and human beings can only acquire the knowledge which is socially constructed and bound to change over time and across cultures.
Egalitarian:	Egalitarianism is a school of thought which is based on the principle that all people are equally important and deserve to be treated equally.
Femininity:	Femininity is a set of qualities or behaviours attributed to the female sex. These qualities and behaviours are constructed by social and cultural norms as well as biological factors. It is different from the female sex in that both females and males can possess feminine traits. Traditionally, feminine attributes include, but not limited to, emotion, nurture and sensitivity.
First Language:	First language refers to the language that people learned first and speak best. It is used especially when a person can speak more than one language.
Foreign Language:	Foreign language learning takes place in contexts where the language being learnt is not easily accessed by learners.
Gender:	Gender is constructed based on social and cultural factors, and generally used to explain the extent to which individuals are masculine and feminine, rather than male and female (i.e., biological sex)
Gender Socialisation:	Gender Socialisation is the process of constructing gender. People learn to behave in a way which corresponds to the socio-cultural expectations of their gender group and develop a gender identity accordingly.
Gender Stereotypes:	Gender stereotypes or stereotyping refer to overgeneralized beliefs about people based on their gender.
Individual Differences:	Individual differences stand for distinctions among individuals on some of the significant psychological traits, personal characteristics, cognitive and emotional components. Among the individual differences come attitudes, values, ideologies,

interests, emotions, capacities, skills, socioeconomic status, gender, height, and so forth.

Language Skills:

Whether it is a second or foreign language, there are four skills which language teachers aim to improve with their language instruction in the classrooms. These are listening, speaking, reading and writing. These four skills are crucial in language teaching and learning as they are key to being thoroughly proficient in an additional language.

Masculinity:

Masculinity is a set of qualities or behaviours attributed to the male sex. These qualities and behaviours are constructed by social and cultural norms as well as biological factors. It is different from the male sex in that both males and females can possess masculine traits. Traditionally, masculine attributes include, but not limited to, aggression, physical strength and confidence.

Mixed Methods

Research:

Mixed Methods Research involves collecting, analysing, and combining qualitative and quantitative approaches at any phase in the research process. The main reasons to adopt mixed methods research in studies might be triangulation, complementarity, development, initiation and expansion.

Mother Tongue:

The language which individuals have grown up speaking from their early childhood rather than a language learned at school or as an adult.

Mplus

Mplus is a latent variable modelling program with a wide variety of analysis capabilities (e.g., exploratory factor analysis, confirmatory factor analysis and structural equation modelling). *Mplus* allows all these modelling features to be combined in a fully integrated general latent variable framework.

Multi-threat

Framework:

The multi-threat framework distinguishes between six distinct stereotype threats arising from the intersection of two dimensions: the target of the threat (the self/one's group) and the source of the threat (the self/outgroup others/ingroup others). These stereotype threats are self-concept threat, group-concept threat, own-reputation threat (outgroup),

	group-reputation threat (outgroup), own-reputation threat (ingroup), group-reputation threat (ingroup).
Native Language:	The language of the country where individuals are born and acquire in early childhood. It is also used interchangeably with 'mother tongue' or 'first language'.
Productive Skills:	The skills which require language learners to produce the language they learn. The productive skills are speaking and writing. They are generally compared with the receptive skills of listening and reading.
Patriarchal:	Patriarchal system is a social structure in which males are believed to be superior to females. Males hold the power and control the society.
Receptive Skills:	The skills which require language learners to receive and understand the language they learn. The receptive skills are listening and reading. They are generally compared with the productive skills of speaking and writing.
Second Language:	Second language learning contexts mostly refer to the contexts where learners have the chance to find the language being learnt used in their immediate environment.
Sex:	Sex is a biological category referring to the biological differences people have. In other words, sex is what people are born with as opposed to gender which is a socially and culturally constructed concept attributed to the biological sex.
Source Language:	In translation, the source language is the one from which a translation is made.
Stereotype Threat:	Stereotype threat is being at risk of confirming, as self-characteristic, a negative stereotype about one's group. In other words, if there is a negative stereotype that is attributed to a group, individuals who belong to that group are at risk of acknowledging it as self-characteristic when the negative stereotype is made salient or applicable.
Target Language:	In translation, the target language is the language being translated to.

CHAPTER 1: INTRODUCTION

1.0. Chapter Outline

The purpose of this chapter is to provide a general introduction to the current thesis. The chapter unfolds in three sections. First, it describes the problem of the study, namely the role of gender stereotyping of academic subjects in academic achievement and gives a brief synopsis of the relevant literature. It further identifies the main gaps in literature and justifies the need for the current research which aims to explore the link between gender stereotyping of foreign language learning and foreign language achievement in the context of Turkish English as a foreign language (EFL) learners. The second section briefly identifies gender roles and expectations of males and females in Turkey and establishes the significance of the current study discussing its context and relevance. Finally, the last section provides an overview of the structure of the thesis and the accompanying chapters.

1.1. Gender stereotyping of Academic Domains

Previous research has largely documented that males and females choose different career paths in early adolescence, which results in gender imbalances in certain academic subjects (van der Vleuten, Jaspers, Maas, & van der Lippe, 2016). For example, while males prefer subjects such as Science, Technology, Engineering and Maths (STEM), females opt for arts and humanities. According to Barone (2011), such 'humanistic–scientific' or 'care–technical' segregation continues to be observed in Higher Education (HE) as well. The gender imbalances in certain academic subjects lead to the biased perceptions of those subjects which, in turn, create gender stereotypes about academic ability that are not completely grounded in females' and males' competencies. Miller, Eagly and Linn (2015), for example, collected data from 350,000 participants across 66 countries and demonstrated that even countries with high overall gender equity (e.g., the Netherlands) held strong gender-science stereotypes if the field of science was dominated by males. However, the data indicated that higher female enrolment in science education in college and above was linked to weaker explicit and implicit national gender-science stereotypes.

According to van der Vleuten et al. (2016), the gender-based choice of fields of study and occupation stems mainly from females' and males' gender ideology. In some societies, it is traditionally believed that males are supposed to be breadwinners, and females are expected to provide caring and nurturing (Barone, 2011; van der Vleuten et al., 2016; Whitehead, 1996). In daily routines, these gender role expectations including

gender stereotyped subject preferences are imposed upon males and females by their parents, teachers and peers (Barone, 2011). Accordingly, some academic routes such as teaching are mainly preferred by females since they are related to traditional caring roles. However, these routes are regarded as low status by males, and they are directed towards some other and better-paid routes in line with their breadwinner role (Barone, 2011). Such gender stereotypes are mostly taken for granted and those who prefer to conform to gender role expectations imposed socially and culturally behave in certain ways in order not to experience the uncertainty that standing out might create (van der Vleuten et al., 2016). Consequently, academic subjects are associated either with femininity or masculinity.

Several investigations have suggested that associating academic subjects with females or males might be problematic in various ways in educational settings. One of the issues is related to perceived self-efficacy beliefs (hereafter, self-efficacy). It has been consistently reported that females tend to have lower self-efficacy in the fields such as STEM, and they are less likely to believe that they possess an innate ability in these male-dominated academic subjects (Beyer, 2014). However, a number of studies have demonstrated that there is a discrepancy between females' self-efficacy and their actual abilities, skills, or performance. That is, they actually perform better than their self-efficacy indicates. According to Beyer (2014), such discrepancy is caused by the fact that the aforementioned academic subjects are commonly considered masculine. As such, females do not feel as comfortable as males in these areas.

Similarly, domain-specific anxiety is also found to be in line with gender stereotyping of academic subjects. For example, according to Hill et al. (2015), studies conducted among secondary students and adult learners have consistently revealed that females report higher maths anxiety compared to males. However, recent research examining the gender differences in math performance has shown that maths anxiety is not related to females' actual maths achievement. Although previous studies indicated that males were more successful at maths than females, it seems that this gap has been closing in recent years. Devine et al (2012), for example, conducted a study among 433 secondary students (165 females; 268 males) and found no gender difference in arithmetic performance, despite females reporting higher maths anxiety than boys.

Although there is ample evidence that gender stereotyping of academic domains is linked to various educational parameters such as self-efficacy and anxiety, it seems that most research has been more concerned with females and their potential underachievement

in so-called masculine subjects such as maths than their 'overachievement' in other fields such as languages which is mostly associated with females (Plante, Theoret, & Favreau, 2009). The current research, therefore, explores the phenomenon of gender stereotyping in the context of foreign language (L2) learning. It is proposed that it is not only females who face the threat of underperforming in such academic subjects as science and maths. There is also a threat for males as they might underperform when it comes to arts and humanities which are traditionally believed to be 'feminine' domains (Carr & Pauwels, 2006; Pomerantz, 2008). Therefore, this study is especially concerned with males and their performance in an academic field which is mostly associated with females, namely L2 learning. Believing that the impact of gender stereotyping of academic domains on achievement is not always a direct one, the current thesis aims to determine the extent to which gender stereotyping of L2 learning is linked to L2 performance via two mediators, namely L2 self-efficacy and L2 anxiety.

1.2. Gender Roles and Expectations in Turkey

It is interesting to study the phenomenon of gender stereotyping of academic subjects in the context of Turkish EFL learners due to the social and cultural factors underlying the gender differences in Turkey. In Turkey, femininities and masculinities are mainly constructed in line with the traditional understanding of gender roles and expectations (Ziya & Koç, 2016). It is expected that males are the main provider for the family while females are homemakers and caregivers (Boratav, Fişek, & Ziya, 2014). This gender segregation is also sustained and legitimised by the religious discourse which has been adopted by the majority of the Muslim population (Boratav et al., 2014). Traditionally, Turkey has been a place where males are regarded as superior to females in many aspects. This is mostly due to the most widely practiced religion in the society, namely Islam which has at least three theological assumptions claiming the superiority of men over women. According to Islam, females are believed to have been created from man's rib, so it is assumed that God's primary creation was man, not woman. Secondly, it is woman that is blamed for expulsion from Garden of Eden which is also known as "Man's Fall". Lastly, it is believed that women are created not only from man but also for man (Kvam, Scheering, & Ziegler, 1999). That is, her existence is just for the sake of man, so it is not of fundamental importance. Even though similar assumptions are made in Judaism or Christianity, the effects of them depend on the extent to which people accept the assumptions and practice them in their everyday practice. According to Caner et al.'s (2016) recent study on gender roles and the education gender gap in Turkey, the traditional gender role expectations still prevail, especially in socially conservative parts of Turkey which results in educational gap between females and males. Some parents,

for example, believe that there is no need to educate their daughters as they are needed for labour or expected to marry while they are willing to educate their boys (Caner et al., 2016). In a family structure, females are expected to be obedient and need to be protected by males whereas males are supposed to look after the family and make final decisions concerning the important matters (Boratav et al., 2014; Ziya & Koç, 2016). As a 'father', males represent authority, discipline and protection. Females, however, have a closer and more emotional connection with their children (Boratav et al., 2014).

According to Boratav et al. (2014), in the context of social change in the country in recent years, gender roles and expectations in Turkey have been adapted in line with egalitarian perspectives (as opposed to the patriarchal structure). That is, masculinity and femininity is not as strict and clear-cut as it was in the past. However, in their research, Boratav et al. (2014) demonstrated that the adaptation of gender roles and expectations were still along the similar lines with the traditional ones. The interviews with fifty married men from different socioeconomic backgrounds (i.e., high, low and middle socioeconomic status) indicated that regardless of their background and education, males tended to possess the traditional masculinity traits, but they negotiated these based on their situation. The below quote from a young university-educated man shows that the traditional gender role division is still prevalent and strongly advocated by Turkish males.

"I come home at night, 'where is dinner, where is tea, where is this, where is that, where are my clothes?' she also takes care of our child, the house, my dinner; life seems to be harder for women, not like how it is for men. (...) Of course there is equality, but the man should still be in the powerful position in this equality since according to our traditional family values, the man is in charge; that's why the man should be a bit more important." (Boratav et al., 2014, p.314)

The research suggested that for males who supported the traditional gender role division were comfortable with their wives being employed outside home as long as they did not disregard their responsibilities at home (e.g., taking care of children and cooking). In light of these findings, it can be hypothesised that Turkish males' and females' career choices might differ from each other as well. While males would prefer jobs with high income, females might choose a career along which they can perform their domestic responsibilities. This hypothesis is supported by recent statistics (2016-2017) provided by Council of Higher Education (2018) in Turkey. According to these statistics, 66% of 55,922 graduates who completed an Education major (e.g., teacher training with subject specialisation) were females. The female dominance is also observed in the majors related to languages (i.e., Language Acquisition and Literature and Linguistics). In Language Acquisition, there were 5,733 graduates, and 68.2% of these graduates were

females. Similarly, from among 11,883 graduates in Literature and Linguistics, 73% of them were females. When it comes to the engineering majors (e.g., civil engineering), however, only 26.5% of 37,061 graduates were females. These statistics, although subtle, suggest that females and males who aim to conform to the common gender roles and expectations in Turkey might make choices based on their gender. As such, it is likely that gender stereotyping of academic subjects is prevalent in Turkey as well.

As of particular interest, the current study aims to determine the extent to which L2 learning is regarded as a female domain by Turkish adult EFL learners (especially by males) for whom learning English is compulsory rather than optional at university level. It also seeks to explore the possible link between their gender stereotyped beliefs and L2 performance via the mediating roles of L2 self-efficacy and L2 anxiety. Previous research has already suggested that in many L2 learning environments, there is a commonly held belief that L2 learning is a feminine domain (Schmenk, 2004; 2007). However, Plante et al. (2009) emphasise that although it is assumed that languages are believed to be a female domain, the actual prevalence of such gender stereotyped belief still needs to be confirmed. Since little research has been conducted to address this limitation, there is a gap in the literature in this respect. It is also important to note that gender stereotyped beliefs and gender stereotyping of academic subjects may not be generalised for all communities since it is highly likely that they change from community to community. Thus, the current research offers a significant contribution to the literature since it provides a fuller understanding of the phenomenon of gender stereotyping of academic subjects by focusing on L2 learning which is regarded as a female domain and exploring it in Turkey where well-defined femininity and masculinity exists.

1.3. The Current Thesis

This thesis is formed of seven chapters including the current introduction chapter, namely **Chapter 1. Chapter 2**, which follows, opens with a detailed description of the common terminologies in Second Language Studies along with the four language skills, namely listening, speaking, reading and writing and explains their relevance to the current research. It further presents a review of the existing literature on gender stereotyping of L2 learning along with L2 anxiety and L2 self-efficacy and establishes the possible link between these three constructs. The chapter concludes with the detailed description of the research aims and questions.

Chapter 3 comprises the methodology used in the current research. It begins by laying out the philosophical foundations on which this study is built. It provides a justification for

the use of a mixed methods research design as well as describing and evaluating the specific details of both quantitative and qualitative approaches taken in this study. It also highlights the ethical considerations of the research.

The next three chapters present the empirical studies. **Chapter 4** is concerned with the first empirical study of the current research. This was based on three self-report questionnaires, examining L2 learners' gender stereotyped beliefs, L2 anxiety, L2 self-efficacy, and the relationship between these and L2 learners' overall L2 performance. As part of this study, a pilot study was also conducted, and all the three questionnaires underwent a rigorous development and validation process. The results gained through these questionnaires in the main study suggested that gender stereotyped beliefs were not linked to L2 learners' performance through their L2 anxiety. The beliefs were mostly associated with L2 learners' self-efficacy, but only for female L2 learners. Males' L2 self-efficacy did not seem to be influenced by any kind of gender stereotyped beliefs tested. These results suggested that self-efficacy was at least one important mechanism by which gender stereotyped beliefs were related to females' L2 performance.

Chapter 5 presents the second empirical study which adopted a qualitative research approach. It aimed to provide in-depth information about the role of language teachers in terms of gender stereotyping of L2 learning. The findings of semi-structured interviews with L2 learners and L2 teachers helped inform the potential sources of gender-stereotyping of language learning to underpin the results gained from Study 1 and Study 3. It was found that even if some L2 teachers held certain gender stereotypical beliefs (e.g., females are more hardworking than males), they did not seem to be related to L2 learners' own gender stereotypical beliefs. The existing gender stereotyped beliefs L2 learners mentioned seemed to be created and legitimised prior to university and mostly by families and the society.

Chapter 6 is concerned with the last empirical study which used an experimental design. It distinguishes between self-as-target and group-as-target stereotype threats and examines the extent to which these exert different effects on males' L2 listening performance. The results showed that there were significant differences between the two experimental groups and control group in terms of L2 performance. The experimental groups performed worse than the control group. It was also tested whether L2 listening anxiety and L2 reception self-efficacy were the mediators of the stereotype threats and L2 listening performance. However, neither L2 listening anxiety nor L2 reception self-

efficacy was found to be the mechanisms that might account for the link between the stereotype threats and L2 listening performance.

The final chapter, **Chapter 7**, provides a general discussion of this research. It integrates and discusses the key findings of all the three studies. It further highlights the theoretical and practical contributions to the literature as well as outlining the limitations and directions for future research.

1.4. Chapter Summary

The current chapter has presented a background summary for the study, outlining the issue of gender stereotyping of academic subjects and its possible causes and consequences in the field of education. In particular, since it would be simplistic to suggest a direct relationship between gender stereotyping of academic subjects and academic performance, the focus has been mostly on the roles of anxiety and self-efficacy which might play a mediating role in the relationship between gender stereotyping of academic subjects and academic performance. The chapter has further explained that the phenomenon of gender stereotyping of academic subjects has been understudied in the field of foreign language learning which is mostly associated with females. Therefore, the current research has been set out to examine the extent to which gender stereotyped beliefs about L2 learning is linked to L2 performance, especially for male EFL learners, through the mediating roles of L2 self-efficacy beliefs and L2 anxiety. Since the research takes place in Turkey, the context of Turkish EFL learners who were the participants in the current study has been introduced as well. Finally, the organisation of the thesis was set out including a brief overview of each chapter.

CHAPTER 2: LITERATURE REVIEW

2.0. Chapter Outline

The aim of this chapter is to provide an overview about the theories and conceptions adopted in the current PhD thesis. The chapter is divided into three main sections. The first section defines the common terminologies used in Second Language Studies (SLS) and elucidates why this differentiation is significantly important to this thesis. The section also identifies the four major language skills, namely reading, writing, listening, and speaking which need to be acquired to be proficient in an additional language (L2). Following this, the second section establishes a link between one of the main individual differences, namely gender and gender stereotyping of L2 learning with a particular focus on stereotype threat (ST). ST, “being at risk of confirming, as self-characteristic, a negative stereotype about one’s group” (Steele & Aronson, 1995, p.797), is discussed in detail with regard to how it might impair language learners’ performance. Finally, the last section sheds light on two of the mediators of ST which are perceived self-efficacy (hereafter, self-efficacy) and anxiety elaborating how these influence the language learning process.

2.1. Common Terminologies in Second Language Acquisition

Language scholars tend to use various terms to explain the theories, processes, contexts and other issues related to learning a language other than the mother tongue (L1). The terms appearing in the literature include, but are not limited to, *second language acquisition* (mostly abbreviated as SLA), *second language studies* (SLS), and *foreign language learning* (FLL). In this first section, a detailed description of these terms is presented highlighting the similarities and differences between them (see Table 2.1 for a summary of the discussion). The terms which are relevant to the current thesis are identified further.

2.1.1. Second Language Acquisition vs. Second Language Studies

SLA is the widely used umbrella term which explains various phenomena related to the processes involved in learning a language other than L1 (Gass & Selinker, 2008; Tavakoli, 2013). As Gass and Selinker (2008) mention, the term does not only refer to ‘second’ languages, but also to third (or more) languages. When it comes to the literal meaning of SLA, however, it is not necessarily accurate because the phrase ‘language acquisition’ might sometimes have a different meaning which is explained in detail below. Hence, to avoid any confusion, some scholars might prefer using the term second language studies (SLS) interchangeably to refer to any study concerned with using or

acquiring a second language (L2) (Gass & Selinker, 2008; Tavakoli, 2013). In line with this, the current thesis adopts the term SLS instead of SLA as the cover term. It is important to note here that the field is widely known by the acronym SLA (Gass & Selinker, 2008). Therefore, the aim here is not to disregard the term SLA completely, but to maintain clarity and consistency throughout the current thesis.

SLS researchers seek answers to the same what, how and why questions pertaining to learning any language in addition to L1. However, the way these questions are addressed differs from each other depending on researchers' approaches to understanding and investigating language. According to Tavakoli (2013), there are three main academic disciplines which inform language research. These are linguistics, sociology and psychology. Linguists' main concern is the nature of language itself. They analyse the characteristics of languages such as phonetics, phonology and morphology. They also examine learners' linguistic competence (underlying knowledge) and linguistic performance (actual production) at various stages of acquisition (Tavakoli, 2013). Sociolinguists focus on variability in learners' actual production which also covers communicative competence. Psychology is a relatively broad discipline, so psychologists and psycholinguists can explore L2 in a number of ways. One is concerned with the mental or cognitive processes involved in language acquisition, and the neural representation of languages in learners' brain. Another way is from the social perspective. Social psychologists investigate how group-related phenomena affect the process of L2 learning. Among these phenomena come identity and social motivation, and the interactional and larger social contexts of learning. Given that SLS draws from and impacts these areas of inquiry, it is indeed an interdisciplinary complex field (Gass & Selinker, 2008). The current PhD project benefits from this interdisciplinary nature of SLS adopting a socio-cognitive approach which draws together individual level factors (i.e., stereotypical beliefs, self-efficacy, anxiety), with environmental and contextual factors (i.e., societal stereotypes and how these are transmitted, socialisation process) with behaviour (i.e., L2 performance).

2.1.2. Second Language Acquisition vs. Second Language Learning

In the field SLS, language *acquisition* is often contrasted with language *learning*. In respect to the difference between these two concepts, the best explanation comes from Krashen whose theory of second language acquisition is widely known and accepted in the field (Lightbown & Spada, 2013). Krashen (1982) suggests that the process of second language acquisition is similar to the process of a child's first language acquisition. Children *acquire* their mother tongue through a subconscious process. As

they are exposed to L1 constantly through natural communication, they are not conscious of the language learning process. Language *learning* is different from language *acquisition* in that people *learn* a language when they deliberately attempt to gain information about the target language (L2). That is, in L2 learning, conscious attention is directed to language forms and rules (Krashen, 1982).

However, not all scholars use these words these terms in precisely the same way. Loewen (2014), for example, indicates that he prefers using acquisition and learning interchangeably while adopting the terms explicit and implicit knowledge to differentiate between conscious and subconscious language learning processes respectively. In a similar vein, second language learning and second language acquisition are used interchangeably in the current thesis unless explicitly indicated otherwise. It is important to note here that while learning and acquisition are not differentiated in the context of 'second' languages, they are treated as different entities when it comes to contrast second language learning and foreign language learning. The reasons for this are discussed in detail in the next section.

2.1.3. Second Language Acquisition vs. Foreign Language Learning

The distinction between second language learning and foreign language learning has also received considerable attention from many language researchers. The basic difference between these concepts is that foreign language learning takes place in contexts where the language being learnt is not easily accessed by learners, whereas second language learning mostly takes place in the contexts where learners have the opportunity to access the language being learnt in their immediate environment (Moeller & Catalano, 2015). As the current thesis focuses on learners in Turkey where English is taught as a foreign language, for the sake of a better understanding of the learning environment, it is important to consider the extent to which these two contexts are different from each other.

Some scholars might believe that there is no need to make such distinction between these two concepts based on the context in which languages are learned. According to Vanpatten and Benati (2015), for example, such distinction is not useful because linguistic, psycholinguistic and cognitive dimensions of language learning do not change across different contexts. They believe that people and the mechanisms used in L2 learning are the same all over the world. However, it needs to be noted that Vanpatten and Benati (2015) do not disregard the effect of context completely. Although they advocate the use of SLA as an umbrella term for all contexts, they accept that context

has an impact on rate of learning and ultimate proficiency and that it has a value from a sociological perspective (Vanpatten & Benati, 2010).

Based on the fact that context indeed matters in language learning when analysed from a sociological perspective, some authors emphasise the importance of acknowledging the differences between learning a second language and learning a foreign language. Gass and Selinker (2008) point out that these two contexts differ from each other in terms of both quality and quantity. In a foreign language learning context, language learners are exposed to the language via the formal classroom instruction. In the classroom, their main sources of input are generally the teacher, the materials (e.g., textbooks) and other learners (Gass & Selinker, 2008).

Table 2.1. Summary of the Common SLS Concepts

Second Language Studies vs. Second Language Acquisition	Second Language Learning vs. Foreign Language Learning	Second language Learning vs. Second Language Acquisition
Second Language Acquisition (mostly abbreviated as SLA) is the umbrella term used to explain various phenomena related to the process of learning another language. However, the literal meaning of SLA is different, so scholars suggest that Second Language Studies (SLS) can be used to prevent any confusion.	Foreign language learning takes place in contexts where the language being learnt is not easily accessed by learners, whereas second language learning contexts mostly refer to the contexts where learners have the chance to find the language being learnt used in their immediate environment.	People <i>acquire</i> a language when they are exposed to language without recognising. Acquisition requires informal language input as well as formal language input. People <i>learn</i> a language when there is a conscious attention to language forms and rules. This happens when there is only formal language instruction available to language learners.

As Gass and Selinker (2008) indicate, while teacher talk might be limited, input coming from learners are both limited and often filled with errors. As such, the quality of input might be questionable in foreign language learning contexts. However, in a second language learning context, learners have the opportunity to receive input from native speakers of L2 outside the classroom. Because they receive both formal and informal language instruction, it can be assumed that in a second language learning environment, learners are in a more advantageous position in terms of quality and quantity of input.

Furthermore, foreign language learners deal with L2 consciously and their focus is mainly on the form (e.g., grammar, syntax etc.) whereas a second language is structured unconsciously, and the main focus is on communication rather than language structures. Given these, the former tends to prioritise knowledge over skill while the latter emphasises skill rather than knowledge.

In Turkey, English is taught as a foreign language. Even though English has become a global language, exposure to English is still limited to the classroom in Turkey since L2 learners do not have enough opportunities to use English outside the classroom (Hamzaoğlu & Koçoğlu, 2016). Clearly, the language learning process is not the same for a language learner (Turkish EFL learners in the current thesis) who is exposed to the target language (L2) in the classroom only for a limited time and a learner who is exposed to L2 constantly in their immediate environment. Although the two contexts might be similar to each other in terms of linguistic, psycholinguistic and cognitive dimensions, they differ from each other in terms of the contextual factors which are acknowledged throughout the current thesis.

2.2. The Four Language Skills

Whether it is a second or foreign language, there are four skills which language teachers aim to improve with their language instruction in the classrooms. These are listening, speaking, reading and writing (Horwitz, Tallon, & Luo, 2010). According to Burns and Siegel (2017), these four skills are crucial in L2 teaching and learning as they are key to being thoroughly proficient in an additional language. In line with this assertion, several international standardised proficiency tests such as Test of English as a Foreign Language (TOEFL) and International English Language Testing System (IELTS) aim to evidence L2 learners' listening, speaking, reading and writing ability. As the current thesis is concerned with all the four skills, the following sections provide an overview of the characteristics of each skill to develop a better understanding what it takes to be a good language learner.

Some language scholars argue that these skills should not be treated discretely because in a real communication, the skills are interconnected to each other (e.g., interacting with people require both listening and speaking skills). As such, separating them from each other does not provide an authentic language learning environment for language learners (Burns & Siegel, 2017). Those who advocate the integrated language skills generally group the skills into two categories, namely receptive skills and productive skills. Receptive skills refer to listening and reading (i.e., learners need to receive and

understand the language). Productive skills consist of speaking and writing (i.e., learners need to produce the language) (Burns & Siegel, 2017). Although recent developments in the field (e.g., communicative language teaching) promote the need for the integration of these skills in teaching practices, it is still a common practice to categorise the core language skills into four major skills in favour of convenience, accessibility and practicality in language teaching and research (Burns & Siegel, 2017). While acknowledging the value of integrating the language skills in language teaching and research, it is important to highlight that each skill is unique in its own way and particular attention needs to be paid to each of them to promote effective language skill development (Burns & Siegel, 2017). Given that, the characteristics of each language skill are explained in detail in the following sections.

2.2.1. Listening

Listening is one of the receptive skills which has not received much attention from language researchers and practitioners until recently (Santos & Graham, 2018). It is only for the last two decades that the key role of listening in L2 performance has been acknowledged in the field of SLS, and a number of researchers have sought to determine the factors playing a role in successful L2 listening comprehension (Santos & Graham, 2018). As there is a growing literature proving the significance of L2 listening in the language learning process, it has now become a crucial part of many language programmes and is delivered as a course in its own right or in combination with speaking which is the other receptive skill (Renandya & Hu, 2018).

Previous research has demonstrated that listening is regarded as one of the most difficult skills to acquire because it poses a number of challenges for L2 learners. Goh (2000) groups the main challenges into two categories: lower level problems and higher level problems. Lower level problems are mostly associated with the nature of the spoken text. These include, but are not limited to, variations in rate of speech and accent, recognising intonation and patterns in sounds and unfamiliar words and grammar structures. Higher level problems are related to learners' failure to grasp the meaning of the spoken text. According to Goh (2000), there are three phases of L2 listening which are perception, parsing and utilisation. While learners experience the lower level problems in the phases of perception and parsing, the higher-level problems occur in the phase of utilisation. In a more recent study, Li and Renandya (2012) explored the sources of listening difficulties using a 38-item questionnaire. Data collected from 301 students and 30 teachers in China showed that there were five groups of factors: spoken text (e.g., vocabulary used in the spoken text), issues with processing (e.g., missing the substantial information),

listener factors (e.g., anxiety), chosen listening tasks (e.g., questions that learners are expected to answer based on the text), and external factors (e.g., the adequacy of listening instruction in the class).

Researchers suggest that listening comprehension problems can be addressed to some extent by an interactive model of listening (McAuliffe & Brooks, 2018). This model requires learners to use both the bottom-up (i.e., paying attention to the smaller components of what is heard such as sounds, syllables and words which are then combined together to form phrases, clauses and sentences) and top down processes (i.e., using prior knowledge about the spoken text such as its topic, genre and participants). According to Santos and Graham (2018), successful listeners are the ones who have the ability to employ the bottom-up and top-down strategies effectively depending on the task demands. Several studies have also investigated the learning strategies that learners might use to enhance their listening comprehension. Among these, Vandergrift (1997) identified three types of strategies, namely metacognitive, cognitive and socio-affective. Metacognitive strategies refer to thinking about how to manage the listening process effectively. These include planning, monitoring and evaluating the listening process. Cognitive strategies require learners to manipulate the language to accomplish the listening task (e.g., guessing the unknown words or note-taking). Socio-affective strategies involve interaction and affective control in language learning (e.g., self-encouragement and reducing anxiety).

Although there is a considerable amount of research contributing to our understanding of the complex nature of L2 listening and offering a number of strategies to address this complexity, listening still remains a challenge in pedagogical practices. According to Renandya and Hu (2018), not all teachers keep their professional knowledge and practice up to date either because they do not have access to the relevant knowledge or they are not able to use the new information appropriately. It is also possible that teaching conditions are not viable to adopt new approaches to teach L2 listening effectively (e.g., lack of access to online resources). As such, most of language teachers continue employing traditional teaching methods which are mostly product oriented as opposed to process oriented. A good example of a product oriented listening teaching activity would be the comprehension-based activities which require students to listen to the listening records a couple of times and answer a series of questions based on their understanding. Such an approach is problematic for two reasons. First, although the number of correct answers might provide clues as to whether learners comprehend the listening record, they do not reveal anything about the process of comprehension

(Vandergrift, 2007). That is, it is not clear whether learners pick the correct answer purely based on their understanding of the listening record or they use some other strategies such as using their prior knowledge. Secondly, when the focus is on the right answer, learners' level of anxiety increases (please see Section 2.3.3.2 for a detailed review of listening anxiety). On the other hand, process-oriented approach to listening (e.g., raising awareness about the listening strategies or providing instruction to aid in word recognition and comprehension) has been proved to be more effective in enhancing learners' L2 listening comprehension (Vandergrift, 2007).

2.2.2. Reading

The other receptive skill which needs to be acquired by L2 learners is reading. According to Burns and Siegel (2017), reading plays a crucial role in L2 learning as a source of language input both inside and outside language classrooms. Grabe (2010) emphasises that it is not simple and straightforward to be proficient in L2 reading because L2 learners need to develop a different set of skills depending on the purposes for L2 reading. The main purposes for L2 reading include, but are not limited to, reading for information (e.g., scanning), reading to learn and analyse and reading for comprehension (Grabe, 2010). When L2 learners read for specific information, they need to be able to scan for a specific word or phrase rapidly. In contrast, reading to learn requires a slower reading rate. L2 learners are expected to make connections between sentences and extract the main idea. Their prior knowledge also plays a role in this type of reading. Reading for comprehension is believed to be the most common type of reading in L2 classes. L2 learners are required to have a good word recognition skill as well as rapid overall reading skill and ability to comprehend the text under time pressure (Grabe, 2010). According to Grabe (2010), a successful L2 reader knows how to use their reading skills effectively to achieve the purpose for their reading.

Given the complex nature of reading, researchers have approached L2 reading from a variety of perspectives and aimed to better understand what it takes to be a good L2 reader. Recent research has revealed that L2 reading comprehension is strongly related to L2 learners' linguistic knowledge. In their recent meta-analysis, Jeon and Yamashita (2014), for example, showed that L2 learners' reading comprehension is highly correlated with L2 grammar knowledge ($r = .85$), L2 vocabulary ($r = .79$), L2 listening comprehension skill ($r = .77$), and L2 decoding skill ($r = .56$). Some studies have also investigated the link between learners' attitudes towards reading in L1 to their attitudes towards L2 reading (Yamashita, 2004; Yamashita, 2007). In one of these studies, Yamashita (2007) discovered that the participants who were anxious in L1 reading

tended to be more anxious in L2 reading compared with those who felt less anxious in L1 reading. According to Murtiningsih and Hapsari (2018), in some contexts, learners are more motivated to acquire other skills such as L2 speaking rather than L2 reading because they think that speaking in L2 is much more important than reading. Additionally, in most language learning environments, L2 reading is heavily based on textbooks which generally feature short reading texts followed by comprehension check questions. As Vraštilová (2018) states, such texts are artificial and fails to interest most L2 learners in language classrooms. It is also believed that they do not stimulate critical thinking among language learners, especially at university level (Murtiningsih & Hapsari, 2018).

Murtiningsih and Hapsari (2018) emphasise that L2 teachers play a crucial role in addressing these issues and assisting L2 learners' reading development. Similarly, Grabe (2011) also calls for an effective and focused reading instruction because L2 learners cannot improve their reading skill simply by reading in L2 a lot. They also need their L2 teachers to enhance their L2 reading skills. As a fundamental step, L2 teachers are expected to choose suitable and interesting reading texts which help increase L2 learners' L2 reading motivation. It is suggested that the chosen texts are neither too easy nor too difficult. They need to be just above students' actual L2 level so that new information can build on their existing knowledge. Furthermore, L2 teachers can design appropriate activities that would assist L2 learners' reading comprehension. For example, before reading any texts, new vocabulary can be introduced to L2 learners or their background knowledge can be activated through some pre-reading activities. As for while-reading activities, L2 learners can be asked to do silent reading, read in groups or read aloud in the classroom. Post reading activities should include inference, prediction or evaluation questions since these promote communication and critical thinking skills among L2 learners (Murtiningsih & Hapsari, 2018).

2.2.3. Speaking

It is believed that speaking in L2 does not only facilitate language acquisition, it also contributes to L2 learners' academic success (Goh & Burns, 2012). Considering the value of L2 speaking, many L2 learners think that success in L2 highly depends on the mastery of the speaking skill and, therefore, aim to be a competent L2 speaker (Goh & Burns, 2012). However, it is not easy and straightforward to achieve this aim due to the complex nature of this production skill (Burns & Siegel, 2017). As Thornbury (2012) highlights, L2 learners experience a number of difficulties while developing their L2 speaking. One of the major constraints is the gap in their knowledge of the target language system which includes grammar, vocabulary and phonology. Since any

insufficiency in these areas might impair language fluency and intelligibility, traditional language instruction mainly focuses on addressing any problems in these areas (Thornbury, 2012). Such an approach is inadequate in terms of improving the competence in L2 speaking because speaking in L2 is more complicated than just acquiring the linguistic knowledge. As mentioned by Gan (2012), some researchers group the functions of speaking into three categories which are interaction (e.g., speaking for communication purposes), transaction (e.g., information exchange) and, performance (e.g., giving a presentation). L2 learners also need to be able to have a good understanding of such genres of discourse and speak in L2 accordingly. Furthermore, as opposed to reading and writing, L2 speaking requires L2 learners to produce the language without having the opportunity to plan and reflect on their performance (Burns & Siegel, 2017). As a result, L2 learners might experience high level of anxiety of other debilitating emotions (see section 2.3.3.2 for a detailed review of speaking anxiety).

To deal with the heavy demands of L2 speaking, L2 learners need to develop certain cognitive (e.g., using an alternative term or using language chunks), metacognitive (e.g., preparing the content and the form of the message) and interactional (e.g., asking for clarification or giving examples to clarify the message) strategies (Goh & Burns, 2012). Previous research has demonstrated that adopting these strategies can enhance L2 learners' speaking performance. Jamshidnejad (2011), for example, demonstrated that the use of communication strategies helped L2 learners speak more accurately and better maintain the flow of conversation. Similarly, Nakatani (2005) conducted an experimental study and investigated the extent to which the strategy training group's speaking test performance differed from the control group's performance. After 12-week EFL course, it was revealed that the participants who received systematic strategy development instruction in the course performed significantly better in the oral proficiency test than the control group.

Although there are several studies showing the benefits of strategy development, they do not necessarily have an impact on the real pedagogical practices concerning L2 speaking (Goh, 2017). As Goh and Burns (2012) observe, even though various speaking activities take place in language classrooms, these do not generally feature any skill or strategy development. L2 speaking classes are mostly seen as the opportunities to practice speaking. However, it is suggested that classroom instruction should aim to provide L2 learners with the opportunities to develop the aforementioned strategies. L2 teachers need to adopt materials facilitating strategy development as well as providing

speaking practice and have the necessary knowledge and skill sets to guide and support their students systematically with their language instruction (Goh, 2017). It is important to note that the development of L2 speaking does not only depend on L2 teachers and students, but also some other external factors as well. According to Tante (2018), for example, in some language learning contexts, speaking constitutes only a small part of the language curriculum. Due to the limited time allocated to L2 speaking activities in class, L2 learners' speaking ability might be premature. It is also possible that L2 teachers themselves might have a limited L2 proficiency which prevent them from offering enough support to their students (Tante, 2018). Considering that L2 learners, especially in EFL contexts, do not have a direct access to L2 outside the classroom, more attention needs to be given to L2 speaking development within the classroom environment.

2.2.4. Writing

As Cumming (2015) states "L2 writing is inherently multi-faceted, involving multiple issues and orientations that may now even be commensurable each other" (p. 65). This has been evidenced by Riazi, Shi and Haggerty's (2018) recent analysis of the empirical research published in the *Journal of Second Language Writing* which is believed to play a key role in determining the content, key topics and research methodologies in the field of L2 writing. The analysis revealed that there are various research areas concerned with L2 writing some of which are feedback, instruction, language and literacy development, assessment, writing strategies, the difference between L1 and L2 writing and L2 writing quality (Riazi et al., 2018). The focus of each of these areas has also been varied as well. The literature on feedback, for example, consists of the studies investigating the role of teacher feedback, peer feedback, error correction in L2 writing development and so on (Riazi et al., 2018).

Although previous research offers valuable insights as to the nature of L2 writing and how to better develop learners' writing skill, several researchers are concerned that most of these studies focus on second language contexts, rather than foreign language contexts (Manchón, 2009). According to Manchón (2009), these two contexts are different from each other in several ways. As mentioned above, in second language learning contexts, learners have the opportunity to interact with native speakers and learn about their culture. However, not all L2 learners in foreign language contexts are provided with such an opportunity, which might have an impact on their proficiency and motivation to develop their L2 writing. Also, it is believed that the ability to write in L2 is quite linked to L2 learners' ability to write in L1. Considering that L2 learners' L1

education differs from each other throughout the world, it is highly likely that their proficiency level in L2 writing is different from each other as well. Furthermore, in second language contexts, L2 writing might be seen as an essential tool to communicate in everyday life while it is mostly neglected in foreign language contexts (Manchón, 2009). Considering these differences, it can be said that L2 writing research should contribute to our understanding of L2 research by approaching both contexts equally and exploring the similarities and differences between them.

These two contexts also differ from each other in terms of the approaches taken to better teach L2 writing. According to Paran (2012), there are three approaches which are mainly discussed by researchers and L2 teachers. These are product, process and genre-based approaches. Product-based approach refers to teaching L2 learners how to mimic a model text. It is mostly concerned with the language use such as grammar, spelling and usage. In contrast, process-oriented approach's main focus is on the writing process itself. It deals with the cognitive process of writing such as drafting and revising based on given feedback. Genre-based approach aims to enhance L2 learners' understanding of why and for whom a text is produced. L2 learners' success in L2 writing highly depends on the teaching approach taken in the classrooms. According to Han and Hiver (2018), for example, genre-based approach might increase L2 learners' self-efficacy and reduce their anxiety as L2 learners are supported throughout L2 writing development process. Hence, it is believed that in many foreign language learning contexts, the common teaching approach is still product-oriented. It is only at university level where genre-based or process-oriented approaches might be adopted.

Additionally, the existing literature suggests that there are some factors which are integral to L2 writing success regardless of whether L2 writing occurs in a foreign or second language learning contexts. According to Polio (2017), these include, but are not limited to, language-related issues such as accuracy and cohesion, knowledge of different genres, processes involved in text production, cognitive and metacognitive strategies and writing goals and motivation. Also, Han and Hiver (2018) highlight the role of psycho-social factors such as self-efficacy and anxiety in L2 writing success (see section 2.3.2.3 for a detailed account of L2 writing self-efficacy and 2.3.3.2 for L2 writing anxiety). Given that, like the other language skills, teaching and learning L2 writing is also a complex activity and requires concentration and effort from both L2 learners and teachers.

The above discussion evidences that the language skills are clearly distinct from each other as they have their own characteristics and dynamics. As such, it is highly likely that L2 learners have strengths and weaknesses in particular skills. While they might be competent in L2 speaking, for example, they might have difficulties in the other skills such as L2 writing. As Dörnyei and Ryan (2015) indicate, even after receiving the same language instruction, there might be major differences among L2 learners in terms of L2 attainment. One of the most important factors playing a role in such variance in L2 learners' success either in a certain skill or overall language proficiency is individual differences which are discussed in detail in the next section.

2.3. The role of Individual Differences in L2 learning

Characteristics of language learners, also referred to as individual differences (IDs) have been widely studied and it is known that these differences are significant contributors to success in learning a language other than the native language (Dörnyei, 2005). Among these IDs come "attitudes, values, ideologies, interests, emotions, capacities, skills, socioeconomic status, gender, height, and so forth" (De Raad, 2000, p.41). Although there are several studies exploring these IDs discretely, there is not enough research explaining how the combination of these particular IDs work in a particular language learning context. As Dörnyei (2005) suggests, there needs to be more combinations of IDs because it provides a better understanding of their contribution to L2 performance. As of particular interest, this PhD project is concerned with the combination of gender stereotyped beliefs (i.e., L2 learning is a female domain), L2 self-efficacy and L2 anxiety. It investigates the link between these constructs and L2 learners' language performance. Therefore, the following sections explore these IDs in detail.

2.3.1. Gender, Gender Stereotypes and L2 learning

Before discussing the role of gender in L2 learning, it is important to underscore the distinction between sex and gender. Although these two terms are sometimes used interchangeably, they have different connotations. While sex is binary and can be easily determined biologically, gender is socially constructed, and, therefore, variable (Mooney & Evans, 2015). As Mooney and Evans (2015) explain, although people are born with a certain sex, namely male or female, the way they construct their identities differ from each other due to the characteristics assigned to their sex by their culture. The differences across genders are generally used to explain the extent to which individuals are masculine and feminine, rather than male and female (Mooney & Evans, 2015). Therefore, it is suggested that when research inquiries aim to explain individuals' behavioural patterns, they should not be concerned with their sex, but their gender which

offers insights as to how people perform their identity and how the society judges their behaviours. Therefore, the main concern in the current research is to investigate foreign language learners' gender, not their sex.

According to Schmenk (2007), exploring the link between gender and L2 learning is not an easy and straightforward task which can sometimes cause researchers to avoid studying in this area. This difficulty might stem from the fact that there are some widespread common-sense beliefs concerning gender and L2 learning, and a great deal of people tend to form these beliefs based on their first-hand experiences rather than empirical evidence (Schmenk, 2007). One of the most common gender stereotypical beliefs in the field of SLS is that of L2 learning being a female domain (Carr & Pauwels, 2006; Pomerantz, 2008; Schmenk, 2004). In one of the earliest investigations into this matter, Carr (2002) interviewed 100 secondary school male students and found that the majority of the participants did not believe that they would succeed in languages. Additionally, they felt alienated from the L2 learning context because most of their L2 teachers were females who tended to favour female learners more and most of the materials and activities seemed to appeal to females mostly. The perceived value of L2 learning as an academic domain was found to be very low among male L2 learners as well. According to the participants, languages did not seem to offer various career options or enough cognitive challenge for them. Carr and Pauwels (2006) extended this study by further interviewing 100 L2 male learners as well as teachers and female L2 learners. The findings echoed those of previous research and also showed that not only males, but also females and teachers believed that languages were more appropriate for females. From among the quotes used in Carr and Pauwels' (2006) study, the below quote clearly reflects the majority of the participants' perception of L2 learning:

"There's a real social pressure. I've had friends come up to me and say: 'why are you bothering to do French? What's the point? When will you use it? Why don't! you just do maths?' I find that girls are really good at it, but I think that's because they're expected to be – they're allowed to be! It seems to be a natural instinct to make distinctions between maths and languages and boys and girls." (Carr & Pauwels, 2006:72).

The phenomenon of L2 learning being a female domain was investigated quantitatively as well. Plante, Theoret and Favreau (2009) recruited 1137 Canadian French speaking students from Grade 6, 8 and 10. The data was gained through a 16-item questionnaire which assessed different aspects of gender stereotypes pertaining to L2 learning. It was found that L2 learning was strongly associated with females by the students of all grades

and both genders. These findings further supported Carr and Pauwels' (2006) earlier work.

In line with these studies, previous research has largely documented that males and females choose different career paths, which results in gender imbalances in certain academic subjects (van der Vleuten et al., 2016). For example, while males prefer subjects such as Science, Technology, Engineering and Maths (STEM), females opt for humanities, arts and social sciences. Some researchers have argued that the gender stereotypic approaches to certain academic subjects are due to the sex differences in brain structure and function which might have an impact in L2 learning (Pearson, 2017). According to such studies, females possess innate ability in learning languages. However, according to van der Vleuten (2016), ability does not fully explain gender stereotypical approaches because there are several studies showing that females' and males' performance did not fit the expected patterns. For example, it has been demonstrated that females outperformed males in the academic subjects which are mostly associated with males (e.g., STEM) (see Stoet & Geary, 2018). Similarly, in a recent study, Wucherer and Reiterer (2016) analysed the data from 64 participants (32 females) and concluded that depending on the type of language task, males' language performance was better than females. While males performed better in the tasks related to pronunciation, females achieved a more desirable outcome in the grammar-related tasks. As such, van der Vleuten (2016) suggests that gender stereotypes are created and reinforced not due to the gender differences in ability per se, but through the gender socialisation process in which it is almost always possible to find some degree of gender role stereotyping.

According to the gender socialisation theories (Fagot, Rodgers, & Leinbach, 2000), individuals are attributed with certain characteristics of femininity and masculinity. Those who prefer to conform to gender role expectations imposed socially and culturally behave in certain ways in order not to experience the uncertainty that standing out might create (van der Vleuten et al., 2016). As it is traditionally believed that males are breadwinners and women are the ones who provide caring and nurturing, males and females choose their career paths accordingly, which results in gender imbalances in certain academic subjects. As a result, academic subjects are labelled as feminine or masculine. As the gender socialisation process is at the heart of the current thesis, what follows is a detailed account of the main gender socialisation theories and how gender socialisation is related to L2 learning.

2.3.1.1. Gender Socialisation Theories

Once the distinction between gender and sex has been established, researchers have shown an increased interest in the process of constructing gender which is called gender socialisation by sociologists and psychologists (Ryle, 2011). There have been many attempts to explain how people learn to behave in a way which corresponds to the socio-cultural expectations of one's gender group and develop a gender identity accordingly. Some of these theories are psychoanalytic theory cognitive-development theory, gender schema theory and social cognitive theory (Ryle, 2011).

- **Psychoanalytic Theory**

Psychoanalytic theory was pioneered by Sigmund Freud in 1895 who investigated how men and women developed their identities, but its application as a gender socialisation theory was later expanded upon in the late 1970s by Nancy Chodorow. As Wharton (2005) states, this theory is different from social learning and cognitive development theories as it suggests that some aspects of gender identity are not learned. They are the result of unconscious psychological processes rather than conscious processes such as observing or creating schemas. As a key factor for the development of gender identity, Chodorow (1978) suggests the role of the mother who cares for her children no matter what sex they are. She argues that children spend more time with their mother than their father, so their first identification begins with someone who has feminine characteristics. In the process of developing their own identities, female children find it easier to develop a sense of identity as they learn how to be a female from their mothers. For male children, however, it is more challenging to develop a male gender identity as they first need to understand that their mother is different from their gender identities and seek for masculine characteristics for themselves to associate with (Stockard, 1999). For male children, being masculine is defined as being 'not feminine'. These children try to separate themselves from feminine characteristics and in this process, male children often learn to devalue femininity as well. It is important to note that neither the classic psychoanalytic theory nor Chodorow's (1978) application of it as a theory have been able to receive any empirical support (Bussey & Bandura, 1999). For example, there was no empirical evidence that suggests mothers bonded with their daughters more strongly than they did with their sons (Sroufe, 1985). Therefore, this theory falls short to answer the question, 'How do people develop their gender identities?'.

- **Cognitive-Development Theory**

Cognitive development theory was developed by Kohlberg (1966) who was influenced by Jean Piaget. Cognitive development theory suggests that children acquire a gender

identity and learn how to behave accordingly in a series of discrete, fixed developmental stages. According to cognitive developmentalists, the first stage of the process in gender socialisation occurs when a two or three-year-old child can identify their own gender as well as identify the gender of others around them. At the age of five, they realize that gender is something stable. Actual gender-typing (which is also known as sex-typing) begins at the age of seven when children achieve gender constancy. They understand that even though the physical appearance changes, gender remains stable. Once they know what gender they are, they feel the need to conform to this role (Bem, 1983). Children start selecting behaviours from their environment which are suitable for their gender identity. Cognitive development theory does not completely dismiss the role of environment in the process of gender socialisation. It also provides those important clues for them to adhere to.

Although there was a lot of interest in Kohlberg's (1966) theory over the decades, similarly to the psychoanalytic theory, the main tenets of this theory also failed to find empirical support. For example, there was no published evidence showing that there was a link between children's attainment of gender constancy and their gender linked conduct (Huston, 1983). It is said that children start playing with toys which are traditionally associated with their gender before they attain gender constancy. This shows that gender constancy is not prerequisite for gender development as suggested by cognitive-development theorists (Bussey & Bandura, 1999). In response to negative findings, cognitive-development theorists attempted to change the gender constancy measure as they believed that it is not the theory itself, but the measure that has failed. However, this attempt could not fully support their argument about gender development, either. Therefore, this theory also could not provide a full answer to how people develop their gender.

- **Gender Schema Theory**

Bem (1983) criticised cognitive development theory because it does not provide an explanation as to why children need to socialise based on their sex group rather than other categories such as race or religion. Bem (1983) argued that cognitive development theory assumes that sex differences are more important to children than other differences because sex is regarded as something natural and inevitable. She noted that it is not enough for children to socialise themselves without any external influence. The role of culture and society cannot be dismissed. There needs to be balance between learning solely by external or internal effects. Bem (1983) suggested that schemas which are a set of cognitive associations enable children to understand the world around them.

According to gender-schema theory, schemas gather and organise information received from the outside world and they share the way people see the world. They change the way people think. Schemas are complex in that they make interconnections between information they gather. In this sense, they enable people to categorise behaviours as masculine and feminine and they also create various associations with those categories (Bem, 1993). According to Bem (1993) gender is an important organising category because almost all societies need gender to function properly. Therefore, children need gender schemas and shape these appropriately.

Gender schema theory is believed to provide a useful framework for understanding the cognitive processing of gender information through gender schemas. However, it also has some limitations to explain gender development. Both cognitive development theory and gender schema theory aim to explain gender conceptions, but they fail to address the ambiguity about the mechanisms by which children acquire gender-linked conceptions and translate them into a gender-linked conduct. Bandura (1986) stated that knowing the gender stereotypes does not necessarily require people to behave in accordance with them. For example, an elderly person who has an idea about negative stereotypes about his/her age does not mean that he/she will adopt the same negative stereotypes (Bussey & Bandura, 1999). These two theories fail to explain variations in gender-linked conduct. However, social cognitive theory (also known as social learning theory) offers possible motivational and self-regulatory mechanisms which underpin gender development and functioning. As the latest model, social cognitive theory, is able to address most of the issues that the other theories have failed to address, the current thesis sticks to social cognitive theory.

- **Social Cognitive Theory**

Social cognitive theory, which informs the current thesis, originates from the legacy of behaviourism. The behaviourists claim that human behaviour is learnt, and they are particularly interested in exactly how people learn these behaviours. According to Skinner, behaviours can be shaped based on a system of punishments and rewards. Human behaviours are sustained by rewards or repelled by punishments. In this theory, gender socialisation works in a similar way (Ryle, 2011). Social learning theorists believe that there are behaviours which are particular to one sex and they name these as 'sex-typed' behaviours (Mischel, 1970). These behaviours can also be categorised as 'sex roles' or 'gender norms' as well (Ryle, 2011). Basically, behaviours are labelled as appropriate for one sex, but not the other. According to social cognitive theorists, children start learning gender when they are guided based on these sex-typed behaviours.

According to this perspective, children are reinforced both positively and negatively for gender appropriate and inappropriate behaviour (Wharton, 2005). For example, if a girl cries, it would be acceptable, and she is not punished, but if a boy cries, he would be told not to do so because crying is a sex-typed behaviour and it is regarded as appropriate for girls, not boys. Through these kinds of interactions, gender socialisation occurs.

It is important to note that the earlier model of social cognitive theory implies that there is a conscious effort in society to punish or reward the sex-typed behaviours. However, in reality, it is not always the case, so social cognitive theorists explain the theory by adding that conscious intent on the part of agents of socialisation is not necessary to the process (Bandura & Walters, 1963). According to Bandura's (1986) social cognitive theory, there are three factors shaping the individual's behaviour: personal, behavioural and environmental influences. Therefore, the idea of observing and modelling the behaviours emerges. According to this argument, the development of children's gender is the consequence of their interaction between the role models around them. They develop their perceptions of gender by observing and modelling the people around them, namely agents of gender socialisation and develop a gender identity for themselves. The following section identifies the main agents of gender socialisation and elaborates how they affect individuals in certain contexts such as schools.

2.3.1.2. Agents of Gender Socialisation

Agents of gender socialisation are groups, institutions or organisations which both formally and informally contribute to individuals' development of gender identity (Bussey & Bandura, 1999). The main agents of gender socialisation are generally identified as the family, the school, peers and mass media (Bussey & Bandura, 1999). According to Wharton (2005), gender socialisation process is twofold. First, children experience and understand the outside world through their parents. As seen above, each gender socialisation theory offers a unique explanation for the process of learning gender, but they all share the idea that people develop a sense of gender identity with the help of their parents. Therefore, the family is seen as the first major agent of gender socialisation (Thompson, Hickey, & Thompson, 2016). As Thompson et al. (2016) highlight, gender socialisation occurs throughout adolescence and into adulthood. During this second phase of socialisation, people continue acquiring cultural and social information from other individuals, groups, and organisations such as the school and mass media. As of particular interest, this section focuses on the second phase of gender socialisation and explores the role of teachers as an agent of socialisation.

Teachers are believed to be one of the most important groups in the educational context (Retelsdorf, Schwartz, & Asbrock, 2015; Stromquist, 2007). As Retelsdorf et al. (2015) put forward, this is because teachers are the ones who communicate with their students every day, teach them and judge their behaviours and performance. Consequently, it is highly likely that they inform their students' cognitive and social development as well as their future career choices. Given their potential influence on students, it is expected that teachers treat female and male students equally in the classroom. However, previous research has revealed that teachers' behaviours towards females and males are not always along similar lines due to their gender stereotyped beliefs (Stromquist, 2007). For example, females have been found to be mostly associated with hard work and diligence while males as assertive, aggressive, competitive and outspoken. As such, males seemed to receive more attention in general from their teachers compared to females (Stromquist, 2007).

Teachers' expectations from males and females in specific subject domains have also been studied by several researchers. Robinson-Cimpian et al. (2014), for example, showed that teachers rated males' performance in maths higher than females. Females were considered successful only on the condition that they were hardworking, behaved better or showed more eagerness towards learning. It is important to note that the studies focusing on teachers' perceptions of their students based on their gender have been mostly concerned with the STEM subjects, rather than the subjects such as languages (e.g., Gunderson, Ramirez, Levine, & Beilock, 2012; Robinson-Cimpian et al., 2014). As such, the literature in this respect is rather limited. Among few studies focusing on languages, Siegle and Reis (1998) demonstrated that teachers' evaluations of males' and females' ability differed from each other. For example, they perceived that females were more successful than males in languages, but they did not think there were any gender differences in other subjects (Siegle & Reis, 1998).

Teachers' such subjective evaluations and expectations are believed to be linked to students' ability related perceptions and their actual performance in certain academic subjects (Gunderson et al., 2012). That is, it is suggested that teachers' implicit or explicit gender stereotyped beliefs and expectations may lead students to perform consistent with these beliefs and expectations. Some scholars label this potential effect as self-fulfilling prophecy (e.g., De Boer, Bosker, & van der Werf, Margaretha, 2010; Robinson-Cimpian et al., 2014). There are some quantitative studies in educational research which indicates that the self-fulfilling prophecy might be true. These studies have shown that

teachers' expectancies and attributions had indeed an impact, although sometimes small, on their students' attitudes and achievement ($r = .10 - .30$) (Robinson-Cimpian et al., 2014). Although these studies offer some insights about the relationship between teachers' gender stereotyped beliefs and their students' beliefs and performance, more research investigating teachers' explicit beliefs about females' and males' different domain-specific abilities is still needed (Retelsdorf et al., 2015). Being informed by the existing literature, in the current thesis, it is hypothesised that L2 teachers might hold some gender stereotyped beliefs about L2 learning and treat females and males differently in the classroom. Consequently, this might be linked to L2 learners' gender stereotyped beliefs, if any. It is believed that such gender stereotyping of academic subjects is problematic because it might have an impact on individuals' performance in certain academic domains. This is referred to as the stereotype threat effect which is discussed in the following section in detail.

2.3.1.3. Stereotype Threat

Research on the effects of stereotypes on performance is not a new phenomenon (Gilbert & Hixon, 1991; Katz, Epps, & Axelson, 1964; Katz, 1964; Katz, Roberts, & Robinson, 1965; Stanley, 1971). However, the first discussions and analyses about it emerged with Steele and Aronson's (1995) research which also used the ST for the first time. According to Steele and Aronson (1995), ST is "being at risk of confirming, as self-characteristic, a negative stereotype about one's group" (p.797). In other words, if there is a negative stereotype that is attributed to a group, individuals who belong to that group are at risk of acknowledging it as self-characteristic when the negative stereotype is made salient or applicable. It is important to note that ST is not limited to social groups who are commonly stigmatised in a society (Steele & Aronson, 1995). Any individual who is a member of a group to which a negative stereotype can be applied is under the risk of being threatened by ST (Steele & Aronson, 1995).

According to ST, widely-known negative stereotypes about one's group result in the immediate situational threat. This threat can be either because of being afraid of being judged and treated stereotypically by the others or of possibly self-fulfilling such a stereotype (Steele & Aronson, 1995). To confirm this theory, Steele and Aronson (1995) conducted four experiments. In this seminal research, they investigated the extent to which a social-psychological predicament namely stereotype threat would affect African American's intellectual test performance. Their assumption was that African American students are the potential targets of prejudice and stereotypes when their performance is being tested via a task that is diagnostic of their ability and competence. The results

gained from the experiments supported their assumption. Study 1 and Study 2 showed that when the African American students were asked to take a test described as diagnostic of ability, the number of correct answers they got in the test was lower than the European-American students. However, when the same test was described as non-diagnostic of ability, no such differences were seen between the groups, so it can be concluded that taking a test that was described as diagnostic was enough to activate stereotypes about African American students' group. Steele and Aronson (1995) also aimed to find direct evidence showing that stereotype activation impairs performance. Therefore, in Study 4, they simply asked African American students to indicate their race before taking a difficult verbal test and consequently, the students performed more poorly on the test than the European-American students.

Table 2.2. Summary of stereotype threat literature examining academic performance

Performance Measures	Sample Population	Stereotype Threat Prime Based on	Research
Maths	Women	Gender	Finnigan & Corker, 2016; Ganley et al., 2013; Jamieson & Harkins, 2012; Johns, Schmader, & Martens, 2005; Keller, 2002; Keller & Dauenhimer, 2003; Kiefer & Sekaquaptewa, 2007; Marx & Roman, 2002; O'Brien & Crandall, 2003; Pennington & Heim, 2016; Picho, Rodriguez, & Finnie, 2013; Ryan & Ryan, 2005; Schmader, 2002; Steele, Spencer, & Aronson, 2002; Stoet & Geary, 2012
STEM (Science, Technology, Engineering, Maths)	Women	Gender	Appel, Kronberger, & Aronson, 2011; Beasley & Fischer, 2012; Cheryan, Siy, Vichayapai, Drury, & Kim, 2011; Deemer, Thoman, Chase, & Smith, 2014; Hill, Corbett, & St Rose, 2010; Shapiro, 2012
Academic Performance	E.g., Asian Women, Latino women	Race and Gender	Armenta, 2010; Gonzales, Blanton, & Williams, 2002; Osborne, 2001; Shih, Pittinsky, & Ambady, 1999
Academic Performance	E.g., African American men/women	Race	Aronson, Fried, & Good, 2002; Brown & Day, 2006; Cohen & Sherman, 2005; Cohen, Garcia, Apfel, & Master, 2006; Cvencek, Nasir, O'Connor, Wischnia, & Meltzoff, 2015; Davis, Aronson, & Salinas, 2006; Scott & Rodriguez, 2015; Smith & Hopkins, 2004

Since its first appearance, this pioneering study has inspired several researchers from different disciplines and it has been replicated with varying demographics in a variety of performance domains including older adults and memory (e.g., Barber & Mather, 2013; Levy, 1996), women and Science, Technology, Engineering, Mathematics (STEM) (e.g., Picho et al., 2013), women and leadership (e.g., Hoyt & Murphy, 2016), white males and athleticism (e.g., Stone, Lynch, Sjomeling, & Darley, 1999). However, as this PhD project investigates the effects of ST on adult male students' foreign language learning performance, my literature review mostly focuses on the ST studies examining this domain, namely academic performance (See Table 2.2).

It is important to note that stereotype threat literature has been widely informed by social identity theory. Social Identity Theory (SIT) is important to ST due to the fact that domain identification is important when exploring the likely effects of ST on specific individuals or groups. That is, if an individual does not identify to the domain under threat, there are unlikely to be any detrimental impacts (Pennington, Heim, Levy, & Larkin, 2016). SIT aims to explain the dynamics of group membership and behaviour (Cameron & Lalonde, 2001). According to this theory, an individual's self-concept constitutes two parts: personal identity and social identity. Social identity is "that part of an individual's self-concept which derives from their knowledge of their membership in a social group (or groups) together with the value or emotional significance attached to that membership" (Tajfel, 1978, p.63). The assumption is that social identity as being part of self-concept play a significant role in sustaining one's self-esteem (Cameron & Lalonde, 2001). The individual always seeks for a positive social identity to maintain self-esteem. This results in social comparisons across the groups all of which aim to be favourable and distinct when compared to relevant out-groups (Cameron & Lalonde, 2001).

In SIT, there are three inter-related mental processes involved in evaluating the "in-group" and the "out-group" (Tajfel & Turner, 1979). One of these processes is called *social categorization*. In this process, individuals are grouped under some categories based on their similarities. Some examples of social categories are black and white, female and male (which are of interest in this project), student, Republican etc. By knowing which group they belong to, people can understand the things about themselves. Individuals who are grouped under certain social categories start behaving in certain ways, which results in *social identification*. For example, being categorised as a male, male individuals will most likely behave within the norms of the group. After individuals are grouped under certain social categories and identified with their groups, they start comparing themselves (in-group) against another group (outgroup). This *social*

comparison may cause some problems such as prejudice, discrimination and stereotyping and SIT aims to address these issues by exploring the role of self-concept in group membership, group processes and intergroup relations (Hogg, 2006). According to SIT, there are three motives that make people compare themselves against individuals in other groups which mostly results in prejudice, discrimination or stereotyping (Wheelan, 2005). These are needed for self-esteem, reduction of uncertainty and desire for optimal distinctiveness.

Although the main hypothesis of all the stereotype threat studies has been grounded on Steele and Aronson's research (1995), the focus and the methods of these studies have not always paralleled it over the last twenty years (Shapiro & Neuberg, 2007). Hence, researchers have needed to either adapt the conceptualisation provided by Steele and Aronson (1995) or introduce a new conceptualisation of ST in line with their studies (Shapiro & Neuberg, 2007). The conceptualisations of ST can be grouped under three categories: those that focus on the self as in the original definition provided by Steele and Aronson (1995); those that focus on the group; and those that combine both (Shapiro & Neuberg, 2007). Researchers focusing on the *self* describe ST as being concerned and worried about substantiating, as a self-characteristic, a negative stereotype about one's group by one's behaviour or performance (Croizet & Claire, 1998; Kray, Thompson, & Galinsky, 2001). That is, "stereotype threat is the resulting sense that one can then be judged or treated in terms of the stereotype or that one might do something that would inadvertently confirm it." (Steele et al., 2002, p.389). In contrast, other researchers emphasise that ST occurs when one's performance is believed to be confirming a negative stereotype about their in-group (Bosson, Haymovitz, & Pintel, 2004; Marx, Stapel, & Muller, 2005; Schmader & Johns, 2003). There are also some researchers who, instead of choosing one approach, blend the self and the group and define ST as a psychological predicament which impairs individuals' performance by revealing the risk of confirming a negative stereotype that applies to their group or to themselves in the eyes of others and in their own eyes (Koenig & Eagly, 2005; Schmader, 2002).

The lack of consistency between the definitions mentioned above has created some ambiguity regarding ST. For example, Aronson et al. (1999) question whether ST is self-threatening because individuals are afraid of being a bad ambassador of their group to the society or because they are simply worried about their own reputation and are afraid of being incompetent in the eyes of other people (Aronson et al., 1999). That is, the concerns about the ambiguity in the nature of the threat have been discussed in the

literature since the emergence of ST. Believing that such ambiguity diminishes the value of ST and prevents it from advancing properly, Shapiro and Neuberg (2007) have proposed a multi-threat framework which is explained in detail below.

- **Multi-Threat Framework (MTF)**

According to Shapiro and Neuberg (2007), there is a lack of clarity in ST literature because ST is mostly believed to be a single construct. However, when the literature is examined in detail, it can be seen that each researcher has a different approach to ST which highlights the variability of ST (Shapiro & Neuberg, 2007). The generalisability of much published research on ST is therefore problematic. Under these circumstances, ST has needed to be fully explained for the sake of the future of ST studies and integration with the extant literature. Therefore, a multi-threat approach has been advocated by some researchers to address such concerns mentioned above (Shapiro & Neuberg, 2007; Wout, Danso, Jackson, & Spencer, 2008)

As a response to this need, Shapiro and Neuberg (2007) have created the multi-threat framework (Figure 2.1 and 2.2). According to this 2x3 framework, there are six qualitatively distinct stereotype threats. That is, stereotype threat can be "...to one's personal self-concept, to one's group-concept, to one's personal reputation in eyes of out-group members, to one's group's reputation in the eyes of out-group members, to one's personal reputation in the eyes of in-group members, and to one's group's reputation in the eyes of in-group members" (Shapiro & Neuberg, 2007, p.112). These threats are categorised under two dimensions – the *target* of the threat (the self and one's group) and the *source* of the threat (the self, out-group others and in-group others). According to Wout et al. (2008), whether self-threats or group-threats are activated depends on the social context. Broadly speaking, self-threats are activated when a negative stereotyped domain is associated with one's personal image. For example, in academic contexts, it is important for individuals to be successful. Therefore, they aim to present themselves as being successful. When there is a negative stereotype that claims otherwise, individuals feel that their self-image is being threatened and this prevents them from showing the ability to project an image of competence (Crocker, Garcia, & Nuer, 2008). Hence, these targeted individuals worry about their personal image in contexts where a negative stereotype can be applied. As a consequence, they perform poorly in academic contexts (Wout et al., 2008). In contrast, group-threats are activated in social contexts that a negative stereotype is associated with one's social group. When there is a negative stereotype targeting the group's image, individuals become worried about potentially being a poor ambassador of their group. Again, this can cause targets

to perform poorly in academic contexts (Wout et al., 2008). Each of the stereotype threats is explained in detail below along with Figure 2.1 and 2.2. In the figures, the targets are depicted as males for the sake of my research. The in-group and out-group members can be changed based on the contexts where stereotype threat occurs.

Self-Concept Threat

This is a self-as-source, self-as-target stereotype threat (Shapiro, 2011). It is defined as the fear that occurs when individuals see themselves as having a negatively stereotyped domain which is normally attributed to their group. For example, Ahmet, a male foreign language learner, might fear that a poor performance in a language test will substantiate the belief in his own mind that he is, by virtue of his gender, less competent than his female classmates. One of the most important factors that affect this threat is the extent to which one believes that the negative stereotype is possibly true, namely stereotype endorsement (Shapiro, 2011). That is, if Ahmet is not convinced that the negative stereotype is possibly true, he would not fear having the negatively stereotyped domain. Therefore, he would not experience self-concept threat.

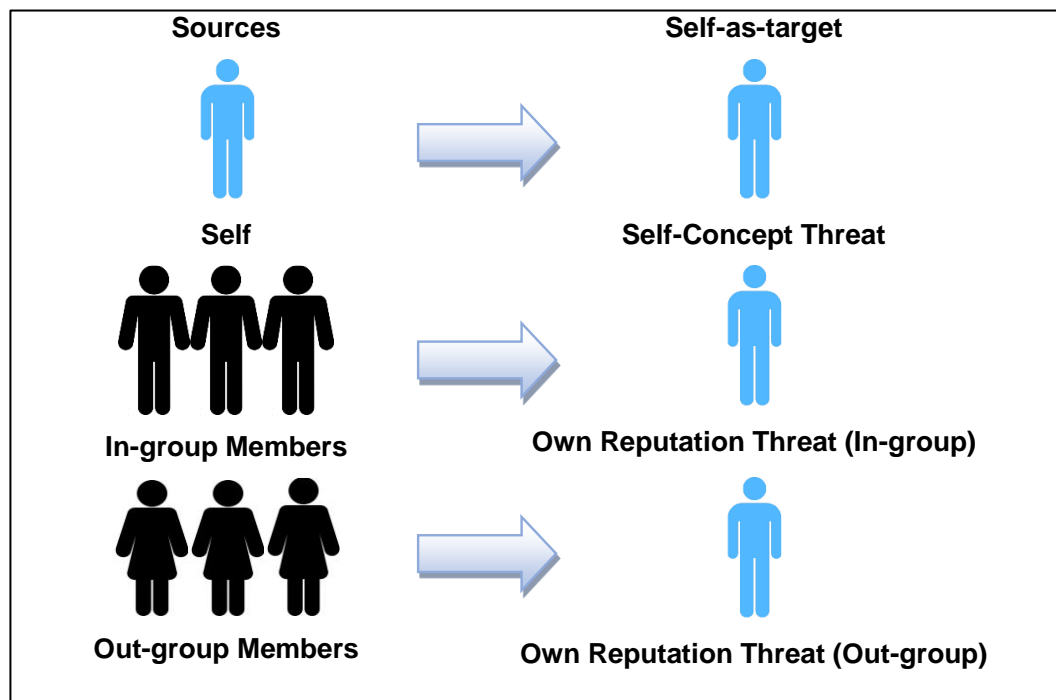


Figure 2.1. *Self-as-target Threats*

Own Reputation Threat (In-group or Out-group)

This is an other-as-source, self-as-target threat (Shapiro, 2011). It occurs when individuals are afraid of being characterised, in the eyes of others, as having the negative stereotype which is attributed to their group. For example, Ahmet might apprehend that

because of his poor performance in an English language test, he would be seen as stereotypic and treated badly by his teachers or friends. To experience Own-Reputation Threat (Out-group), Ahmet should be afraid of being judged or treated badly by an out-group member such as a female friend or a female English teacher. As for Own-Reputation Threat (In-group), the person who judges Ahmet's performance must be an in-group member such as a male friend or a male English teacher. Own-Reputation Threats are different from Self-Concept threat in that unlike self-concept threat, these threats are experienced when individuals believe that others who judge their performance endorse the negative stereotypes about their group (Shapiro, 2011). That is, it is not required for Ahmet to believe that the negative stereotype about his group is possibly true. If he believes that his teacher (male/female) endorses the negative stereotype, he would experience one of the Own-Reputation Threats.

Group-Concept Threat

This is a self-as-source, group-as-target threat (Shapiro, 2011). It occurs when individuals are afraid of legitimatising the negative stereotype about their own group in their own mind by having the negative stereotypic characteristic themselves. That is, they experience group-concept threat when they believe that they are a bad ambassador of their group. For example, Ahmet, a male foreign language learner, might fear that a poor performance in a language test will substantiate the hypothesis in his own mind that male students are, by virtue of their gender, less competent than female students. One of the most important factors that affect this threat is the extent to which one identifies himself with the group, namely group identification (Shapiro, 2011). That is, if being a male is not central to Ahmet's self-concept, he would not be afraid of being a bad ambassador. Therefore, he would not experience group-concept threat. Similar to self-concept threat, it is again necessary for individuals to believe that the negative stereotype is possibly true.

Group Reputation Threat (Out-group-In-group)

This is an other-as-source, group-as-target stereotype threat (Shapiro, 2011). When individuals are afraid of substantiating the negative stereotypes about their group in the eyes of out-group and in-group members, they experience group-reputation threat. For example, Ahmet might apprehend that because of his poor performance in an English language test, his group would be seen as stereotypic and treated badly by his teachers or friends. It is necessary for individuals to believe that the out-group or in-group members believe the negative stereotypes about their group are possibly true. They also need to identify themselves with the group as in the other group-related threats because

those who highly identify themselves with the group feel more pressure to disconfirm the negative stereotype in order not to be a bad ambassador of the group (Schmader, 2002).

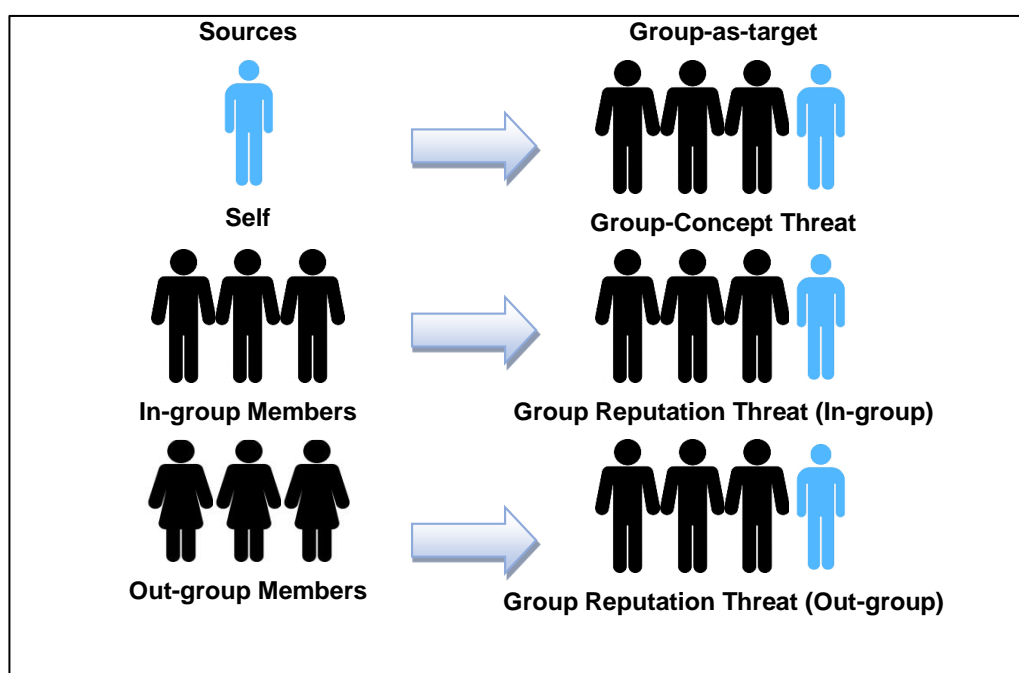


Figure 2.2. *Group-as-target Threats*

Although promising, the multi-threat framework has not yet received enough empirical data revealing the distinct processes behind each of the distinct stereotype threats (Pennington et al., 2016; Shapiro, Williams, & Hambarchyan, 2013). The current study aims to address the gap in the literature by focusing on two of the target of the threats which are self-concept threat and group-concept threat. The reason for this is that no research to date has explored the multi-threat framework in respect of male language performance, and thus it is aimed to contribute to the literature by applying the relevant aspects of this framework to L2 learning and providing initial evidence (for further details about the experimental part, please see Chapter 6).

- **Susceptibility to Stereotype Threat**

Most studies have shown that since there are a number of mechanisms through which stereotype threat causes performance decrements, the way individuals are affected by ST differs from each other (Pennington et al., 2016). In their recent systematic literature review, Pennington et al. (2016) identified a variety of different affective (e.g., self-efficacy and anxiety), cognitive (e.g., working memory and mind-wandering) and motivational processes (e.g., effort and motivation) that might explain the effects of stereotype threat on performance. As of particular interest, the next sections focus on

self-efficacy and anxiety and explain the extent to which these can mediate the effects of gender stereotyping of L2 learning.

2.3.2. L2 Self-Efficacy

According to Mills (2014), research in L2 self-efficacy has become increasingly significant in SLS in the past two decades. This is because the focus of studies in SLS has been shifted from teacher-centred to learner-centred approaches in recent years (Nosratinia, Saveiy, & Zaker, 2014; Riazi, 2007). That is, learners and learning have become more prominent in SLS research. There is now growing evidence confirming that self-efficacy is linked to L2 performance in various ways which are discussed in detail below. Given that self-efficacy is a key asset to L2 learning, it is necessary to study it among L2 learners at different levels of language proficiency and from different perspectives. In the current thesis, L2 self-efficacy is investigated in relation with L2 learners' gender stereotyped beliefs about L2 learning and their L2 anxiety. The next sections analyse both L2 anxiety and L2 self-efficacy at length and examine the extent to which these two important concepts are linked to learners' gender stereotypical beliefs. Before discussing the role of self-efficacy in L2 learning in detail, it is important to conceptualise clearly what self-efficacy is. Therefore, in the section that follows, the concept of self-efficacy is defined.

2.3.2.1. Conceptualisation of Self-Efficacy

Self-efficacy is the central component of the social cognitive theory which posits that the way individuals behave is influenced by the interplay between personal, environmental and behavioural factors (Bandura, 1997). Self-efficacy is defined as people's beliefs or judgments about their own ability to succeed in a variety of specific tasks or situations (Bandura, 1997; Bandura, 2006). People's self-efficacy beliefs can control their functioning through cognitive, motivational, affective, and decisional processes (Bandura, 1997). As Bandura (1997) states, self-efficacy beliefs are so influential that although people have similar knowledge and skills, whether they succeed or fail depends on their levels self-efficacy. Self-efficacy does not only affect whether people think in self-enhancing or self-debilitating ways, but it also predicts how well they motivate themselves and how they react when they are faced with any difficulties (Bandura, 1997).

Schunk and Pajares (2009) point out that self-efficacy enhances human achievement through a number of ways including task choice, effort and persistence. For example, people with stronger self-efficacy tend to invest more time and effort in a particular activity. Also, they demonstrate more perseverance when faced with challenges since

they perceive challenges as opportunities to master rather than threats to their accomplishment and wellbeing. In case of failures, those who have stronger self-efficacy can emerge resiliently from disappointment (Schunk & Pajares, 2009). In contrast, people with lower self-efficacy might perceive some tasks to be more difficult than they actually are. As such, they become more anxious and stressed, which prevents them from thriving in certain tasks and activities (Schunk & Pajares, 2009).

However, it is important to note that even though self-efficacy is deemed to be an important concept, it should not be regarded as the sole influence on behaviour (Bandura, 1997). As Schunk (2012) emphasises, it is a prerequisite that individuals possess the necessary skills and competencies for self-efficacy to be effective. That is, if individuals do not have the skills and competencies needed to accomplish a task, no amount of self-efficacy can produce any desirable results. Task value and outcome expectations also play an important role in helping self-efficacy affect individuals' achievement (Schunk, 2012). For example, learners who feel highly efficacious in L2 learning need to value L2 learning and put some effort in learning a new language. Their high self-efficacy can only then lead to stronger L2 performance. Similarly, learners need to believe that their actions will result in desirable outcomes. As Schunk (2012) mention, for example, if learners believe that their hard work and diligence will not be appreciated by their teachers, they might decide not to engage in particular tasks or activities. In such cases, self-efficacy cannot have a positive impact on academic achievement per se.

According to Mercer and Williams (2014), the concept of the 'self' has gained a particular interest by researchers in the field of SLS in recent years. The increased interest in self-related concepts has resulted in some confusions about theoretical conceptualisations and overlapping terms (Marsh et al., 2018; Mercer & Williams, 2014). Such confusion is a threat to a measure's content validity which refers to the extent to which a measure accurately covers all facets of a construct (Cohen, Manion, & Morrison, 2013). Failing to differentiate between these concepts results in mis-measurement issues which makes it difficult to properly assess the effects of L2 self-efficacy on students' language learning performance (Mills, 2014). Gorsuch (2009), for example, proposed that self-efficacy consists of both cognitive and affective components and measured the affective self-efficacy with items such as 'I am willing to participate in roleplays about daily life in the L2' which in fact measures students' L2 motivation. Similarly, Magogwe and Oliver (2007) used the Morgan-Jinks Student Efficacy Scale developed by Jinks and Morgan (1999) to explore the relationship between Botswana students' language learning strategies and self-efficacy beliefs. This scale consisted of items such as 'It does not

matter if I do well in school' which measures perceived value of L2 learning and 'I am a good L2 student' which measures L2 self-concept rather than self-efficacy. As these scales combine the items that represent different constructs, they lack content validity.

Therefore, to ensure content validity of self-efficacy measures and increase the quality of research, it is crucial to differentiate self-efficacy from the other self-related concepts such as self-concept and self-esteem (Bandura, 2006). According to Marsh et al. (2018) self-efficacy is a 'pure' set of judgements about one's ability to successfully perform a task whereas self-concept is influenced by social comparison and comparable judgements to others. Therefore, measures of self-concept might have self-efficacy items, but measures of self-efficacy have to be context-specific (Zimmerman & Cleary, 2006). Self-esteem is quite different from self-efficacy in that it is an affective reaction showing the extent to which a person value himself or herself whereas self-efficacy is cognitive (Zimmerman & Cleary, 2006). Additionally, it is suggested that items in a self-efficacy measure are phrased in 'can do' which refers to judgment of capability rather than 'will do' which shows intention (Bandura, 2006). Although self-efficacy is believed to be the main source of intention, they are different from each other both conceptually and empirically (Bandura, 2006).

2.3.2.2. Measuring Self-efficacy

According to Zimmerman, Schunk and DiBenedetto (2017), there are five criteria which differentiate self-efficacy measures from other self-belief measures. The first criterion is related to the type of self-belief being assessed. Self-efficacy measures assess individuals' cognitive judgements of their capability to perform a specific task or activity whereas other self-belief measures deal with affective feelings of self-worth or more general judgments of personal adequacy and competence. The second criterion concerns how individuals evaluate themselves. Self-efficacy measures ask them to evaluate their abilities based on a goal-mastery standard. However, most of the other self-belief measures ask them to compare their skills and abilities to others (e.g., I can't learn English as well as my female peers). The third criterion refers to the predictive nature of self-efficacy measures. individuals need to be asked to complete these measures prior to any actual performance. Self-efficacy measures also allow researchers to attain individuals' self-evaluation at varying levels of specificity, which is the fourth criterion. For example, language learners might evaluate their abilities at language skill level or course level. However, other self-belief measures aim to predict individuals' performance in a more global way (e.g., English course), rather than paying attention to specific task features or contextual issues. The last criterion is concerned

with changing nature of self-efficacy. Since individuals' reactions to their experiences might change over time and across different situations, self-efficacy measures cannot provide trait-like individual differences which are stable. Since self-efficacy beliefs are sensitive to many other factors, it is advised that self-efficacy measures can be used more than once to capture any changes over time (Zimmerman et al., 2017).

In line with the above conceptualisation of self-efficacy, Bandura (2006) emphasises that self-efficacy measures need to be specific to particular domains or tasks. Measures that assess generalised beliefs about students' abilities are not predictive as they force students to evaluate their competence without a clear task in mind (Bandura, 2006; Mills, 2014; Pajares, 1996). Some studies have developed measures in L2 self-efficacy literature that have addressed the need for context specificity (Mills, Pajares, & Herron, 2006; Mills & Peron, 2009; Wang, Kim, Bong, & Ahn, 2013). However, the examples of such scales that are designed to assess L2 self-efficacy pertaining to specific L2 learning skills are limited, so more research needs to be conducted in this field. For example, Mills et al. (2006) developed a French Self-efficacy scale which aimed to assess students' reading and listening self-efficacy. There were 14 items for reading self-efficacy and 21 items for listening self-efficacy which were mainly based on the guidelines of American Council on the Teaching of Foreign Languages (1986). Similarly, Wang et al. (2013) created a Questionnaire of English Self-Efficacy (QESE) which was the first attempt to design a scale that assesses all the L2 learning skills, namely speaking, reading, writing and listening. Originally created with interviews and observations, this 32-item scale was tested using Rasch Rating Scale Model on a sample of 167 EFL students in Korea. The major problem with this measure was the lack of gradations of challenge (Wang et al., 2013; Wang, Kim, Bai, & Hu, 2014). According to Bandura (2006), items should represent a mixture of easy and difficult tasks in order to avoid ceiling and floor effects. If there are no obstacles to overcome, all students would be highly efficacious which is not a reliable result, so self-efficacy should be evaluated against varying skill levels (Bandura, 2006). As Wang et al. (2013) state the QESE does not provide a variety of easy and difficult items. As such, a need has arisen to develop a new measure for the purpose of the current thesis. The development and validation of the new measure, the *Questionnaire of Self-Efficacy Beliefs in L2 learning*, is provided in detail in Chapter 4.

2.3.2.3. The Role of Self-Efficacy in L2 Learning

Previous studies have demonstrated that L2 self-efficacy is linked to L2 learners' performance in various ways. For example, learners with higher self-efficacy beliefs are

reported to achieve higher L2 performance scores (Hsieh & Kang, 2010); have lower L2 anxiety (Mills et al., 2006) and use L2 learning strategies more effectively (Graham, 2007; Graham & Macaro, 2008). Also, while students with low L2 self-efficacy may tend to spend more time on simple and straightforward tasks with minimal effort and patience, students with higher self-efficacy beliefs tend to be more willing and put more effort when it comes to challenging tasks (Mills et al., 2006).

The link between L2 self-efficacy and the specific L2 learning skills which are listening, speaking, reading and writing have also been studied by several researchers. In one of these studies, Mills et al. (2006) conducted a study with 95 students learning French at intermediate level (independent user) and multiple regression analysis revealed that the students who reported higher self-efficacy regarding reading received higher scores in French reading proficiency. Within the same study, it was found that French listening self-efficacy was positively correlated with French listening proficiency only for the female students (Mills et al., 2006). Mills and Peron (2009) investigated the effect of a global simulation course which requires students to work collaboratively on a specific task on the students' writing self-efficacy beliefs. Results showed that students who reported higher French writing self-efficacy beliefs had lower writing anxiety, and higher writing self-concept, perceived value of writing, and self-efficacy for self-regulation. As for reading, Li and Wang (2010) carried out an empirical study with 139 Chinese EFL students to explore the link between reading self-efficacy and the use of reading strategies. The study showed that reading self-efficacy was significantly and positively correlated with the overall use of reading strategies. That is, the students with high self-efficacy used more reading strategies compared to those with low self-efficacy. Compared to the other skills, speaking is the least studied skill in L2 self-efficacy literature (Asakereh & Dehghannezhad, 2015; Liu, 2013). In a recent study, Asakereh and Dehghannezhad (2015) explored the link between student satisfaction with the EFL speaking classes, speaking self-efficacy and speaking achievement. The findings gained from 100 Iranian EFL learners at university revealed that there is a significant positive correlation between student speaking self-efficacy and speaking achievement. Overall, these results indicate that L2 self-efficacy is indeed crucial to improve the language learning skills of learners.

Given that self-efficacy is a key asset to L2 achievement, it is necessary to develop an understanding of the link between gender stereotyping of L2 learning and L2 learners' self-efficacy and achievement. It is hypothesised that gender stereotyping of academic domains is problematic because how an academic subject is perceived by females and

males might influence their achievement related perceptions such as self-efficacy (Pajares & Schunk, 2001). In alignment with this argument, several studies have found that females tend to have higher self-efficacy beliefs in language studies compared to males (see Wang et al., 2014). This raises the concern that such differences in self-efficacy might result in discrepancy in L2 achievement between males and females. Although the mediating role of self-efficacy between gender stereotyped beliefs and academic achievement has been studied by researchers concerned with the so-called masculine subjects such as maths, the issue has received far too little attention in the field of SLS (Plante et al., 2009). Thus, one of the aims of the current study is to address this limitation by providing an account of the issue of gender stereotyping in relation to L2 learning.

2.3.2.4. Sources of Self-Efficacy

Bandura (1997) identifies four main sources of self-efficacy which are mastery experiences, vicarious experiences, verbal persuasion and physiological and emotional arousal. Before clarifying each source of self-efficacy, it is important to note that these sources do not affect individuals' self-efficacy automatically (Schunk & Pajares, 2009). The formation of higher or lower self-efficacy based on these sources depend on how individuals interpret their experiences (Schunk & Pajares, 2009). It is also possible that individuals rely on one or a combination of two or more sources of self-efficacy to form their self-efficacy.

From among four main sources of self-efficacy, *mastery experiences* are regarded as the most influential source of self-efficacy. Since they are individuals' own direct experiences, they provide the most authentic evidence whether they can succeed in a task. While self-efficacy is increased by individuals' successes, any failures can undermine it. Bandura (1997) emphasises the importance of experiencing some difficulties and setbacks as well to acquire a resilient sense of efficacy. When people achieve successes easily, they might assume that it will be case in any future tasks, and therefore, can lose their courage when faced with failure. In contrast, they learn how to deal with failures after gaining some experience in overcoming obstacles. Difficulties provide people with opportunities to explore their strengths and use them to turn failures into successes. Given that, establishing high L2 self-efficacy through mastery experiences highly depends on how L2 learners interpret or judge their own successes and failures. For example, getting high grades in language proficiency exams might be a good indicator of success for most L2 learners, especially in foreign language contexts where L2 performance is mainly assessed through exams.

Individuals' self-efficacy might be linked to *vicarious experiences* as well. People might have a role model in their mind and associate themselves with that person. In such a case, their level of self-efficacy changes based on their role model's performance. If the role model is successful, people's self-efficacy is enhanced, whereas if the role model's performance is poor, this might lower their self-efficacy. Vicarious experiences are particularly important for those who do not have any direct knowledge of their capabilities and need some kind of indicators of success and failure. In L2 learning process, L2 learners' role model might be their language teachers. The way L2 learners interpret their teachers' behaviours and compare themselves to these would determine their L2 self-efficacy. For example, having a successful male teacher might increase the level of male L2 learners. In case of any failures, it might help them consider that being successful in L2 learning, as a male, is not impossible.

Verbal persuasion refers to any specific feedback on individuals' performance or a talk which is given to motivate people. Even though verbal persuasion is not believed to create long-lasting effect on efficacy, it has the potential to help individuals make an effort or try hard enough to succeed. However, as Bandura (1997) indicates, individuals should be encouraged within realistic bounds only. That is, verbal persuasion should not raise any unrealistic beliefs, which might result in individuals' failure. This does not only discredit the persuaders, but also weaken individuals' sense of self-efficacy further. Additionally, the effectiveness of verbal persuasion also depends on the credibility, trustworthiness and expertise of the persuader (Bandura, 1986). According to Schunk and DiBenedetto (2016), for example, when capable students doubt themselves and impede their success, a well-respected teacher's verbal persuasion plays a significant role in helping these students clear their self-doubt and thrive. As mentioned earlier, in foreign language contexts, the main source of information is language teachers. As such, in the case of L2 learning, L2 learners might appreciate any encouragement received from their language teachers in or outside the classroom.

It is possible that people use their *physiological and emotional states* as indicators of their self-efficacy. When individuals feel anxious or worried during a particular task, their self-efficacy would be low. In contrast, when they are excited about the task, their self-efficacy would be high (Bandura, 1997). Social cognitive theory proposes it is also believed that the link between anxiety and self-efficacy beliefs may also be the other way around. It suggests that anxiety both affects and is affected by self-efficacy beliefs, so it can be said that there is a bidirectional relationship between self-efficacy and anxiety

(Bandura, 1986; Bandura, 1997). As MacIntyre (1999) highlights, there is a negative correlation between L2 anxiety and both L2 learners' performance and the way they rate their ability in L2 learning. Learners' weakened sense of self efficacy makes them become anxious about their performance in the academic subjects in question. Evidence of this has been found by MacIntyre and Noels (1994). Their study showed that students with high levels of anxiety underestimated their ability in language learning. As L2 anxiety is linked to L2 learners' self-efficacy and plays a crucial role in determining L2 performance and achievement along with it, it is important to enhance our understanding of this construct as well. Therefore, the following section provides a detailed account of L2 anxiety.

2.3.3. L2 Anxiety

The debilitating effect of L2 anxiety on L2 learning and achievement has long been discussed by many different scholars. Krashen's (1982) widely known affective filter hypothesis posited that some affective variables such as motivation, self-esteem and anxiety might impede or facilitate language learning process. That is, L2 learners with low motivation, low self-esteem and high level of anxiety cannot progress in language learning because these variables raise learners' affective filter which forms a mental block and prevents learners from receiving language input effectively. Arnold and Brown (1999) also emphasise that L2 anxiety is possibly one of the most destructive factors in the process of L2 learning. These claims have been supported by plenty of empirical evidence gained through studies done in different countries around the world with different kinds of language learners (for a recent meta-analysis, see Teimouri, Goetze, & Plonsky, 2019).

2.3.3.1. L2 Anxiety in Three Phases

According to MacIntyre (2017), there have been three main approaches to studying L2 anxiety - the confounded approach, the specialised approach and the dynamic approach. The confounded approach refers to the initial studies back in the 1970s. During this period, the studies investigating L2 anxiety revealed conflicting results (e.g., Chastain, 1975; Kleinmann, 1977). While some studies reported that there was a negative correlation between L2 anxiety and achievement, several others could not find a link between anxiety and language learning (Horwitz & Young, 1991; Horwitz, 2017; MacIntyre, 2017; Scovel, 1978). There were also studies showing a positive correlation between L2 anxiety and performance which suggested the idea of facilitating anxiety (Horwitz & Young, 1991; Horwitz, 2017; MacIntyre, 2017; Scovel, 1978). The confounded phase was ended when Scovel (1978) argued that the mixed and confusing results of

previous L2 anxiety studies might be due to using existing anxiety scales to assess L2 anxiety. He called for a specialised approach and suggested scholars to consider L2 anxiety as a complex and dynamic psychological construct and conceptualise it precisely before attempting to measure it with any other anxiety scales (Scovel, 1978).

Scovel's (1978) review is regarded as the turning point in L2 anxiety literature as thereafter it has been a priority for researchers to find the best approach to this psychological construct and measure it appropriately (Horwitz, 2010; MacIntyre, 2017). The earliest rigorous attempt to address the gap that Scovel (1978) mentioned was made by Horwitz et al. (1986). According to Dewaele, (2017), the phase of the specialised approach started with their publication. In this seminal paper, not only did Horwitz et al. (1986) define L2 anxiety as "distinct complex of self-perceptions, beliefs, feelings, and behaviours related to classroom language learning arising from the uniqueness of the language learning process", but they also created the Foreign Language Classroom Anxiety Scale (FLCAS) which has generated a great deal of interest in language anxiety literature and been treated as the best L2 anxiety measure ever since. Indeed, it is still being used in various contemporary studies with different groups of language learners including China (e.g, Jin, de Bot, & Keijzer, 2017), Spain (e.g., Martin & Valdivia, 2017; Santos, Gorter, & Cenoz, 2017), Taiwan (e.g., Kao, Chen, & Craigie, 2017), Pakistan (e.g., Gopang, Ansari, Kulsoom, & Laghari, 2017) and Turkey (e.g., Elaldi, 2016).

However, a more dynamic approach which is more exploratory has been called upon by several researchers (Dörnyei & Ryan, 2015; Pavlenko, 2013). As Pavlenko (2013) states, psychological factors such as motivation, anxiety and willingness to communicate are not one-dimensional. These are continuously interacting with many other interpersonal, social and environmental factors (MacIntyre, 2017). In line with these suggestions, the past decade has seen the rapid development of a more contextualised way of studying L2 anxiety (Dewaele, 2013; Dewaele & MacIntyre, 2014; Gregersen, MacIntyre, & Meza, 2014; MacIntyre, 2012; MacIntyre & Serroul, 2015). Following the current trend in L2 anxiety literature, this study aims to investigate the extent to which L2 anxiety plays the role of a mediator between gender stereotyping of L2 learning and L2 achievement.

2.3.3.2. Skill-based Approach to L2 Anxiety

Although the FLCAS is still popular among the majority of the L2 anxiety researchers, some researchers raise the concern that the items in the FLCAS mainly focus on the speaking skill. This has been evidenced by a recent study conducted by Park (2014).

Believing that many of the studies exploring the dimensions of the FLCAS were disadvantaged by using exploratory factor analysis (EFA) which is conducive to subjective labelling of factors, Park (2014) followed a two-step process in analysing the components of the FLCAS and performed both an EFA on a sample of 217 L2 learners at a university in Korea and a CFA on a different sample of 244 participants from the same university. The EFA revealed two components: Communication Apprehension and Understanding, and Communication Apprehension and Confidence. The goodness of fit indices (CFI = 0.899 and IFI = 0.900) obtained from the CFA showed that two-factor solution for the FLCAS is reasonable. However, as these components were highly correlated with each other, it was concluded that the core component of the FLCAS might be *Communication Apprehension* only. This finding is consistent with those of Panayides and Walker (2013) who tested the reliability and validity of the data gathered from a sample of 304 senior high school EFL students in Cyprus using the FLCAS. The Rasch model they adopted confirmed the unidimensionality of the FLCAS. As there is little evidence confirming that the FLCAS is a combination of a number of factors, it can be concluded that the FLCAS is in fact unidimensional.

The dominance of the speaking-related items in the FLCAS might stem from the belief that speaking is the most anxiety provoking skill among L2 learners (Cheng, Horwitz, & Schallert, 1999). According to Gkonou, (2014), unlike the other three skills, speaking makes the students' self-esteem potentially vulnerable as it involves a higher level of self-exposure. As such, L2 anxiety is mostly associated with speaking and several investigations into L2 anxiety have been concerned with it (Gkonou, 2014; Mak, 2011; Woodrow, 2006). Recent studies, however, have emphasised that it is not only speaking, but also the other three skills (i.e., listening, reading and writing) that need to be studied in L2 anxiety literature as well (Cheng, 2017; Kim, 2002; Luo, 2012). As Horwitz (2017) mentions, some learners might find listening, reading and writing more anxiety-provoking than speaking. Taking this into account, researchers who are interested in a skill-based approach to L2 anxiety have made several attempts to assess the language skills adopting alternative approaches which are discussed in detail below.

Speaking Anxiety

Speaking is the skill most associated with L2 anxiety. This is mainly because it is believed to be the most anxiety-provoking skill (Cheng et al., 1999). Unlike the other three skills, speaking makes the students' self-esteem potentially vulnerable as it involves a higher level of self-exposure (Gkonou, 2014). Therefore, most of the L2 anxiety studies mainly focus on speaking anxiety. To measure speaking anxiety of students studying English in

Australia, Woodrow (2006) created Second Language Speaking Anxiety Scale. It consisted of twelve five-point Likert type items. Woodrow (2006) proposed that for the target population, speaking anxiety can be divided into two categories: in-class anxiety and out-of-class anxiety. The results indicated that although in-class anxiety and out-of-class anxiety are highly correlated, they are two different constructs and the Second Language Speaking Anxiety Scale is valid and reliable to measure them. Woodrow (2006) also reported significant negative correlations between both in-class and out-of-class anxiety and learners' oral performance which is consistent with the notion of debilitating language anxiety (Aida, 1994; Horwitz, 2017; Mak, 2011). In another study, Gkonou (2014) used the FLCAS to investigate the components of in-class speaking anxiety. It was confirmed that speaking anxiety is a source of L2 anxiety. It was also found that fear of negative evaluation and low self-perceptions of speaking ability increase the level of speaking anxiety (Gkonou, 2014).

Listening Anxiety

It is not only speaking but also listening that generates anxiety for language learners. However, the literature on listening anxiety is much more limited compared to speaking anxiety (Chang, 2008; Elkhafaifi, 2005; Zhang, 2013). According to Vogely (1998), it is crucial to detect anxiety pertaining to listening because learners need to comprehend what is being said in order to communicate in L2 effectively. She further suggests that there are two important steps to take in researching listening anxiety. First, the effects of listening anxiety need to be revealed. Second, the sources of it are to be explored and possible solutions should be presented accordingly. Elkhafaifi (2005) adapting Saito et al.'s (1999) Foreign Language Reading Anxiety Scale found evidence supporting that L2 anxiety which was measured by the FLCAS and listening anxiety are different but related constructs. The findings of this study also confirmed that both L2 anxiety and listening anxiety affect language learning achievement negatively. Some researchers who investigate listening anxiety have preferred to create new scales rather than adopting the FLCAS which is primarily concerned with speaking. The Foreign Language Listening Anxiety Scale was one of the earliest attempts to measure language learners' listening anxiety (Kim, 2000; Kim, 2005). The scale consisted of 33 items which formed into two factors: tension and worry over listening and lack of self-confidence in listening. This study also reported that there is a negative relationship between L2 listening anxiety and learners' listening proficiency at university level in Korea. As for the sources of listening anxiety, Oxford (1993) stated that students feel anxious during listening tasks when the tasks are beyond their ability or when students believe that they need to hear every single word. In a qualitative study, Chang (2008) interviewed 140 participants and summarised

the results in four categories. The interviews suggested that there are four major sources of listening anxiety: 1) characteristics of L2 input (e.g. lack of clarity of difficulty, 2) processing related aspects of L2 (e.g. lack of processing time), 3) instructional factors (e.g. lack of practice) and 4) characteristics of teachers or learners (e.g. fear of failure).

Reading Anxiety

Compared to the skills associated with oral communication (i.e. speaking and listening), reading is much less studied in L2 anxiety literature. Because reading (if it is silent) is an individual activity, it is believed to create less trouble for learners. During reading, learners are not susceptible to the factors that might create anxiety (e.g. fear of evaluation) as such. However, contrary to the expectations, reading is also anxiety provoking (Saito et al., 1999; Sellers, 2000; Zhao, Guo, & Dynia, 2013). Saito et al.'s (1999) study offered some important insights into reading anxiety. Not only did they propose that reading makes learners anxious as well as speaking and listening do, but they also created the Foreign Language Reading Anxiety Scale which has been utilised by many researchers later on (e.g., Zhao et al., 2013; Zin, 2010). Using their own scale, Saito et al. (1999) also confirmed that L2 reading anxiety is a distinct phenomenon, but it is related to L2 anxiety in general reading anxiety and it has a negative impact on reading performance. According to Saito et al. (1999), there might be two reasons why reading elicits anxiety: 1) unfamiliar scripts and writing systems and 2) unfamiliar cultural material. In the first instance, learners try to decode the given text. For this, they need to have a good understanding of sound-symbol correspondences in the target language. If they fail to process the text, it is highly likely that they feel anxious. Although learners manage to associate the sounds with the words, the text may not necessarily make sense for them due to being unfamiliar with the cultural material. This ambiguity creates reading anxiety as well.

Writing Anxiety

As is the case with reading, writing is believed to be far less anxiety provoking than the oral communication skills (Cheng, 2004a). That's why, it is difficult to find extensive research on this construct and report on its effects on L2 writing performance. To measure L2 writing anxiety, early studies mainly adapted Daly and Miller's (1975) Writing Apprehension Test (WAT). However, this measure was originally designed for first language (L1) learners and most of the studies adopted this measure without validating it in an L2 context (Cornwell & McKay, 1999). Cheng's (2004b) multi-dimensional Second Language Writing Anxiety Inventory (SLWAI) was the first measure designed exclusively to assess learners' writing anxiety in L2. It is different from many other skill-based scales

in the literature in that it consists of three dimensions which are somatic anxiety, cognitive anxiety and avoidance behaviour. Cheng's study (2004b) suggested that the scale was valid and reliable. Using the SWAI, Tsai and Cheng (2009) found that the lower the learners' writing anxiety, the better they performed in the writing task, so it was concluded that writing anxiety has a significant effect on writing performance. However, more studies need to be warranted to be able to generalise this finding. As for the sources of writing anxiety, Cheng (2004a) conducted a mixed methods study and categorised her findings into four main areas. Learners tend to have writing anxiety due to factors relating to instructional practices (e.g. unreasonable time constraints), personal beliefs about writing and learning to write (e.g. desire for making no mistakes), lack of self-confidence (e.g. believing that they are not good enough), and interpersonal threats (e.g. fear of negative evaluation).

2.3.3.3. Test Anxiety

Generally speaking, test anxiety is caused by fear of failure (Horwitz et al., 1986; Meijer, 2001; Tsui, 1996). It has two major components: worry which is associated with the cognitive concern over failure such as having negative expectations and emotionality which refers to physiological reactions experienced during tests such as sweating (Liebert & Morris, 1967). According to Putwain (2008), test anxiety is situation or context specific. That is, it is important what students are being asked to do. It is also a social dimension as it depends on how students' performance is being judged or evaluated by other people (Putwain, 2008). Test anxiety is regarded as a significant factor affecting learners' L2 performance (Gregersen & Horwitz, 2002; Horwitz et al., 1986; Joy, 2013). However, because test anxiety is widely considered not to be specific to L2 learning (Aida, 1994; MacIntyre & Gardner, 1989), there is not a precise definition or conceptualisation of test anxiety pertaining to L2 learning. That being said, there is little research on L2 test anxiety and how it affects learners' L2 performance and there is not a proper valid and reliable measure specifically designed to assess L2 test anxiety (In'nami, 2006). To date, most research investigating L2 test anxiety have used either the FLCAS which has only 3 items representing L2 test anxiety or other general test anxiety scales (e.g., Sarason & Mandler, 1952). According to In'nami (2006), this kind of approach to measuring L2 test anxiety is problematic for two reasons. First, the validation procedures for the existing measures are not thoroughly implemented. Second, there is not enough number of items in the measures being used to assess L2 anxiety (In'nami, 2006).

As there are problems with the measurement of L2 anxiety, those few studies which have attempted to study L2 test anxiety and its effect on L2 performance produced unclear results. For example, using Sarason's (1961) Test Anxiety Scale, Chastain (1975) showed that there is a negative correlation between learners' performance in an audiolingual French course and their level of test anxiety. However, in the same study, it was revealed that there is not a relationship between performance and anxiety in regular French or German courses. It was also found that there is a positive correlation between test anxiety and learners' grades in Spanish (Chastain, 1975). Using the FLCAS, Horwitz et al. (1986) found that there is a negative but weak relationship between the course grades of the students learning English at beginner level and their L2 test anxiety. To have consistent results regarding L2 test anxiety, the issue of measurement needs to be solved and more research needs to be done on the relationship between L2 anxiety and learners' L2 performance.

2.3.3.4. Measuring L2 Anxiety

Some of the aforementioned studies have managed to produce some valid and reliable measures that can be used to assess L2 anxiety pertaining to specific language skills. However, these scales are not convenient as such when research needs to be done on reading anxiety, writing anxiety, listening anxiety and speaking anxiety simultaneously which is the case in the current study. The main reason why using the validated scales is not feasible and practical is that bringing all the scales together will create a lengthy and cumbersome measure (Cheng, 2017). The number of the items in the skill-based anxiety scales range from 12 (e.g., the Second Language Speaking Anxiety Scale (Woodrow, 2006)) to 33 (e.g., the Foreign Language Listening Anxiety Scale (Kim, 2005)). A questionnaire made of four validated scales would consist of about 70-80 items, which is not either feasible or practical. Shortening the existing validating scales is also challenging because it creates reliability and validity issues. Widaman et al. (2011) note that short forms of measures might have less reliability because of having fewer items and less validity because of biased selection of items from original versions. Also, as MacIntyre (2017) states, research into L2 anxiety has moved towards a contextualised dynamic approach. That is, researchers are interested in studying L2 anxiety and its interaction with many other factors such as perceived self-efficacy and willingness to communicate (MacIntyre, 2017). Therefore, using a scale that has 70-80 items is not ideal for researchers as they need to use other measures as well. Thus, it is evident that there is a need for a new brief measure which enables researchers to assess all the four skills simultaneously. To date, apart from Cheng's (2017) more recent four brief measures (i.e., L2 speaking anxiety scale, L2 reading anxiety scale, L2 writing anxiety

scale, L2 listening anxiety scale), such measure is not available in L2 anxiety literature. Also, since the interest has been shifted towards the specific language learning skills, measuring L2 test anxiety has been majorly overlooked. Most research investigating L2 test anxiety use general test anxiety scales rather than devising a new measure which is specific to language learning. In response to this need, Study 1 produced and validated a new multi-dimensional language class anxiety scale (MLCAS) which helps assess learners' anxiety regarding all the language skills. Differently from Cheng's (2017) measures, the MLCAS assesses L2 test anxiety as well which is also a factor that is highly associated with L2 anxiety (Joy, 2013). The development and validation procedure can be found in Chapter 4.

2.4. Devising the Initial Research Questions

Based on the literature review above, three research questions have been devised. The questions are presented in detail below:

1. Is there a relationship between L2 learners' existing gender stereotyped beliefs about L2 learning and their L2 self-efficacy, L2 anxiety and L2 performance? (Study 1 & 3)

It is hypothesised in the current research that L2 performance is related to several other factors (see Figure 2.3). The assumption is that L2 learners are affected by the society they are in. This effect comes through gender socialisation, in which it is almost always possible to find some degree of gender role stereotyping (Martin, 1990). Learners who are attributed with characteristics of femininity and masculinity start believing that they need to conform to these characteristics and behave accordingly. As discussed in the literature review, such behaviours might result in gender stereotyping of academic subjects (e.g., maths is for males and languages are for females).

Several studies have provided empirical evidence showing the phenomenon of some academic subjects being stereotyped might be true. In one of these studies, for example, Whitehead (1996) found out that female students had negative attitudes towards science and it was less likely for them to choose it and be successful in it. This was associated with the belief that STEM subjects are regarded as masculine and arts and languages are believed to be feminine (Whitehead, 1996). This kind of beliefs is believed to be one of the results of gender stereotyping in general. As Whitehead (1996) notes, contrary to popular myth among people, it is not biological differences in ability that account for differences in the perception of academic subjects and achievement. It is these gender-stereotyped beliefs that belong to a particular aspect of society (Whitehead, 1996). The

impact of gender stereotyping of academic domains on academic performance is not necessarily a direct one. It can be through the mediating roles of other factors such as self-efficacy beliefs and anxiety. Therefore, the current research first aims to explore whether gender stereotyping of L2 learning occurs in the context of Turkish male EFL learners and investigates the extent to which such stereotyping is linked to their L2 performance through the mediating roles of L2 self-efficacy and L2 anxiety.

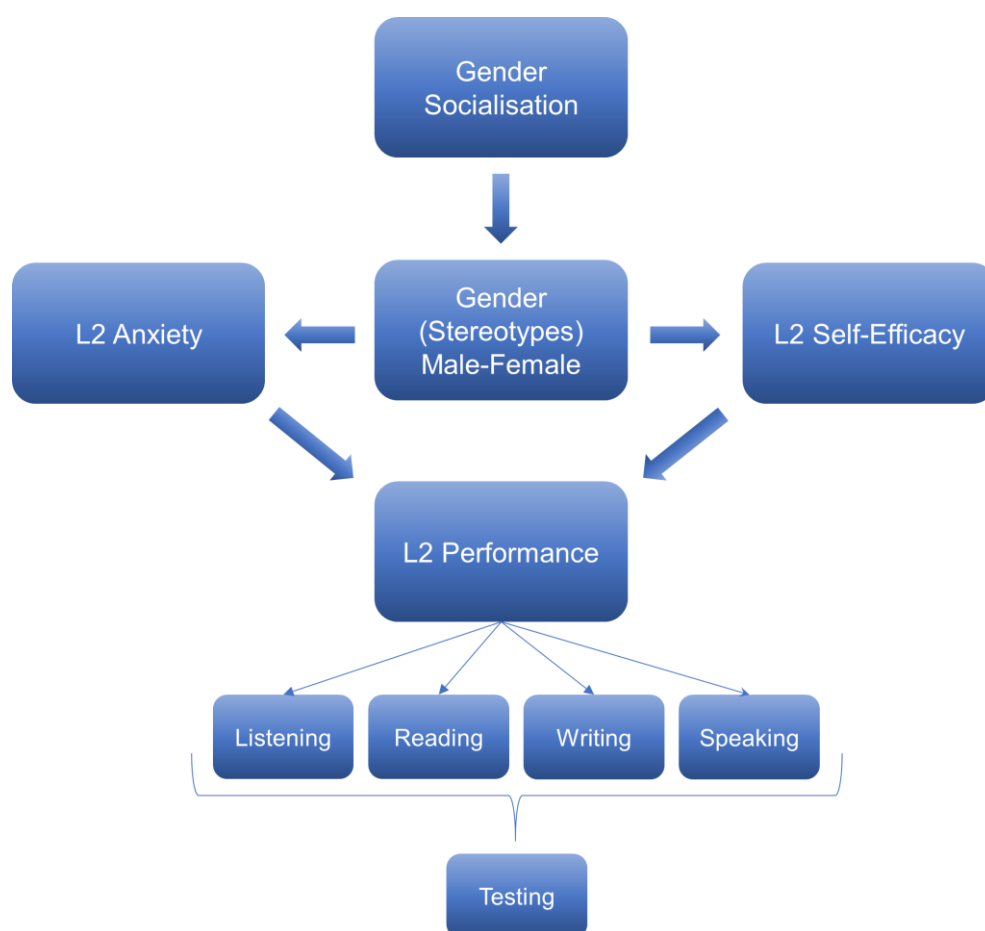


Figure 2.3. *Conceptual overview of the programme of research*

2. To what extent does the learning context play a role in sustaining and legitimising gender stereotyped beliefs in L2 learning? (Study 2)

According to Bussey and Bandura (1999), there are some agents who are responsible for gender socialisation. The main agents of gender socialisation are parents, peers, media, and school (Bussey & Bandura, 1999). After the family, the school and teachers are believed to be significant in terms of developing gender orientations. Therefore, the current research aims to determine the extent to which teachers, as an agent of socialisation, play a role in sustaining or legitimising L2 learners' gender stereotyped beliefs, if any.

3. To what extent does the target (self-concept threat vs group-concept threat) of gender stereotyped beliefs impact upon learners' L2 performance via self-efficacy and anxiety? (Study 3)

Question 3 aims to complement Research Question (RQ) 1. RQ1 seeks to find a link between learners' gender stereotyped beliefs and how they affect their L2 performance through the mediating roles of self-efficacy beliefs and anxiety in general. However, it has some limitations in that it does not demonstrate whether there is a confounding variable affecting the relationship between the endorsement level of the students and their performance. There might be several other factors affecting these two domains (e.g., socioeconomic status, students' motivation and interest etc.). Therefore, to make the research more rigorous, the aforementioned assertions are tested in an experimental design. As the nature of experimental studies suggests, it is possible to control all the other factors, but the targeted one that might affect the dependent variable (Coolican, 2014). Informed by stereotype threat literature, this question, therefore, aims to experimentally elicit whether self-concept threat or group-concept threat differentially affect L2 performance (Figure 2.4).

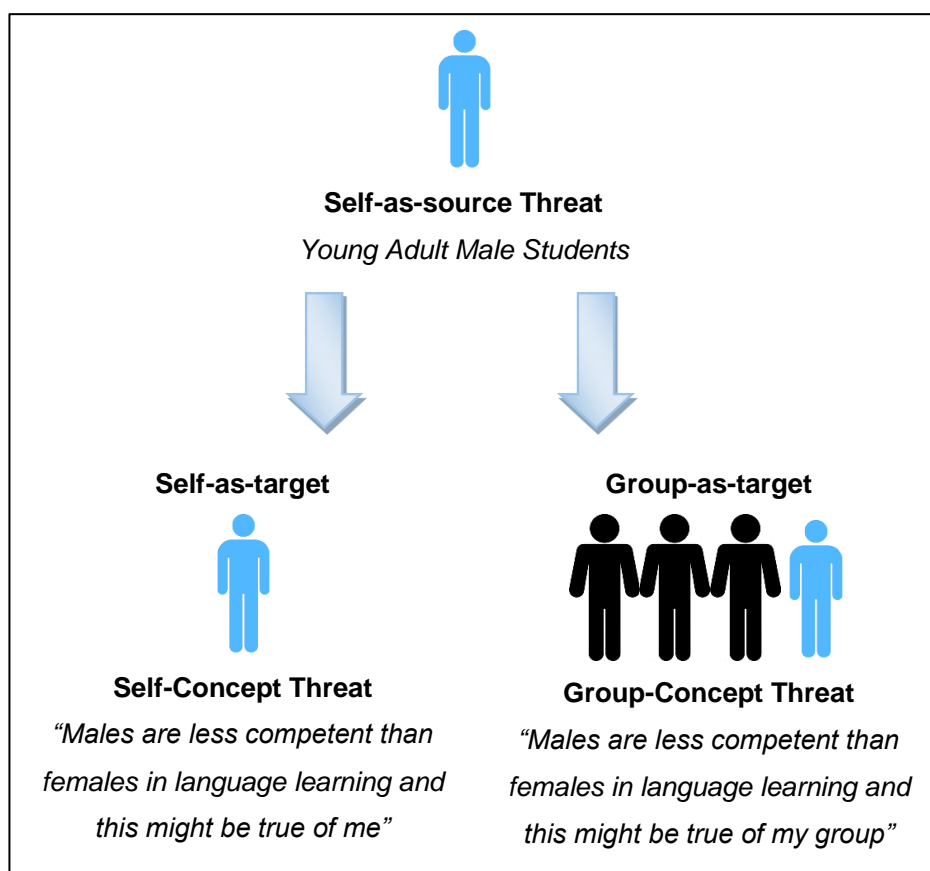


Figure 2.4. Self-concept threat vs. Group-concept threat

Also, Shapiro et al. (2013) point out that most research so far has characterised stereotype threat as a function of either the self or the group as the target. Therefore, in their research, they provide initial evidence for the processes that explain group-as-target and self-as-target threats better (Shapiro et al., 2013). They investigate how stereotype threat interventions namely in-group role models and self-affirmations moderate the effects of ST on African American students' performance. It was found that self-affirmation interventions were effective in reducing the ST effects only against self-as-target stereotype threats, not against group-as-target threats (Shapiro et al., 2013). This study aims to extend their research and inform the findings of previous research in the literature by investigating how self-efficacy and anxiety mediate ST.

2.5. Chapter Summary

Chapter 2 has provided a comprehensive review of literature drawn upon in the current thesis. The chapter has begun by clarifying the distinction between the common terms and concepts most relevant to L2 learning. This has been followed by a description of each language skill (i.e., listening, speaking, writing and reading) which needs to be acquired to be good at a foreign language. The rest of the chapter has reviewed the literature on three main concepts, namely, gender stereotypes in L2 learning, L2 anxiety and L2 self-efficacy. The review of the literature has logically progressed onto the final part of the chapter which has been concerned with the initial research aims and questions of the current research.

CHAPTER 3: METHODOLOGY

3.0. Chapter Outline

This chapter outlines the research methodology and design that were undertaken in the current PhD project to address the research questions. It commences with the description of the underlying research paradigm and the methodological approaches it entails. This is followed by a rationale of the mixed methods research (MMR) approach, along with a review of the particular qualitative and quantitative methods used (i.e., semi-structured interviews, questionnaires and experiments). The data collection instruments, the data collection and analysis procedures and the participants involved in the current research are considered in detail as well. The chapter concludes with the ethical considerations and the limitations pertaining to the methodological choices.

3.1. Critical Realism as a Research Paradigm

Paradigms have recently gained increasing attention from researchers as it is suggested that each researcher should explicitly discuss their paradigmatic foundations that guide their research practice (Creswell & Clark, 2017; Shannon-Baker, 2016). According to Blaikie and Priest (2017), the research paradigm is an unavoidable part of social research and it concerns what needs to be studied, what questions are crucial to address regarding this subject, how it needs to be studied and how the results need to be interpreted. Although its importance in research is well-established, there seems to be a disagreement and inconsistency among researchers regarding what 'paradigm' means and how it contributes to research (Shannon-Baker, 2016). Therefore, before discussing the philosophical assumptions that underpin the current research, it is important to elucidate the concept of paradigm and explore its role and significance in research. Considering that one's choice of paradigm constitutes the basis of their research, this clarification is important and necessary.

The term 'paradigm' was first introduced and popularized by Kuhn (1970). As Masterman (1970) indicated, Kuhn's definition and use of paradigm was not a precise one because there were twenty-one definitions available in Kuhn's (1962) work which consequently created an inevitable ambiguity about the meaning of the term. Hence, Freshwater and Cahill (2013) emphasise that, as paradigm is an "elusive" concept, researchers need to be cautious when approaching it. Following Freshwater and Cahill's (2013) suggestions, different conceptualisations of paradigms discussed in the literature are treated as a continuum (from the least generalisable to the most generalisable) rather than as a

paradigm-or-not dichotomy in the current project. What this means is explained in greater detail below.

According to Morgan (2007), there are four versions of the paradigm concept discussed in the literature all of which are interrelated (Figure 3.1). The first and most specific version of paradigms is as model examples of research which was mentioned and favoured by Kuhn (1962). This version of paradigms is utilised to show new researchers how a particular field deals with its central issues (Morgan, 2007). Paradigms as shared beliefs were also discussed and favoured by Kuhn (1970). This version features shared beliefs within a group of researchers who have a mutual understanding of what needs to be studied and how this should be done (Morgan, 2007). Although these versions are the roots of the concept of paradigm, these have received little attention within discussions in social sciences research and the examples of them are relatively few compared to the other two versions discussed in detail below (Morgan, 2007).

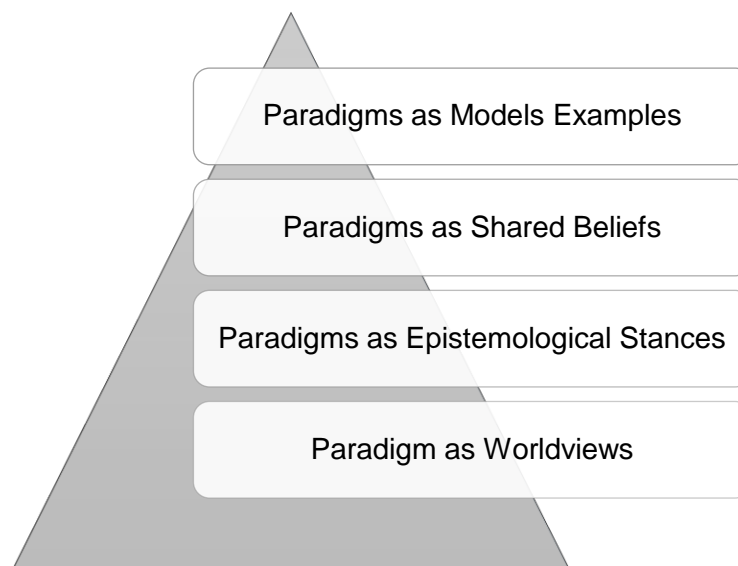


Figure 3.1. *Four Versions of Paradigms*

Adapted from Freshwater and Cahill (2013)

Paradigms as worldviews is the broadest version which covers all the ways of experiencing and thinking about the world (Morgan, 2007). It also acknowledges the role of personal experience and culture in research. Some scholars in social research follow this pattern and equate paradigms with worldviews (e.g., Creswell & Clark, 2017; Guba & Lincoln, 1994). According to Morgan (2007), taking paradigms as a worldview is not always useful in informing research practice because it is too broad. It is necessary to establish the boundaries around 'worldview' rather than aiming to include researchers'

thoughts and beliefs about everything (Morgan, 2007). Paradigm as an epistemological stance, which is believed to be dominant in social sciences, has a narrower focus as it focuses on researchers' views about the nature of research and features distinctive belief systems (e.g., realism and positivism) that provide researchers with guidance on how research questions are asked and answered (Morgan, 2007). Since each belief system has its own understanding of the nature of knowledge, the ways the questions such as "What constitutes knowledge?" and "How is it gained?" are answered differ from each other depending on the epistemological stance taken.

As each researcher approaches paradigms from a different angle and how it is used depends on the purpose of the user, it is difficult to come up with a conventional definition which would address the full complexity of the concept of paradigm. The use of paradigms in the current study is consistent with that of Blaikie and Priest (2017) who adopts Kuhn's views (1970) and define paradigms as shared beliefs which comprises:

“...views of the *nature of reality* (ontological assumptions), concepts theories and *techniques of investigation* that are regarded as appropriate (epistemology), and *examples* of previous scientific achievements that provide models (exemplars) for scientific practice”

As Blaikie and Priest (2017) emphasise, ontological (the study of being, becoming, existence, or reality) and epistemological (the study of knowledge and knowing) assumptions are the key components of a research paradigm. These are followed by two other major concepts, namely methodology and methods (Crotty, 1998) (Figure 3.2).

Given the complex nature of paradigms, it can be noted that the discussions around 'quantitative' and 'qualitative' being synonymous with paradigms are proven to be futile. These are 'approaches' that guide researchers on how they need to deal with data and methods (Shannon-Baker, 2016). As Bergman (2010) states, these tell researchers little about the nature of reality and how they need to approach it. Therefore, these two are not part of the discussion in this section. Except for these, there are various research paradigms commonly adopted in social sciences. Among the popular options come positivism/post-positivism, constructivism, transformative-emancipation, pragmatism and critical realism (Creswell & Clark, 2017; Hall, 2013; Shannon-Baker, 2016). In mixed methods research, however, which is concerned in the current study, the paradigm issue is much more complicated (Greene & Hall, 2010), and the available options are rather limited and controversial.

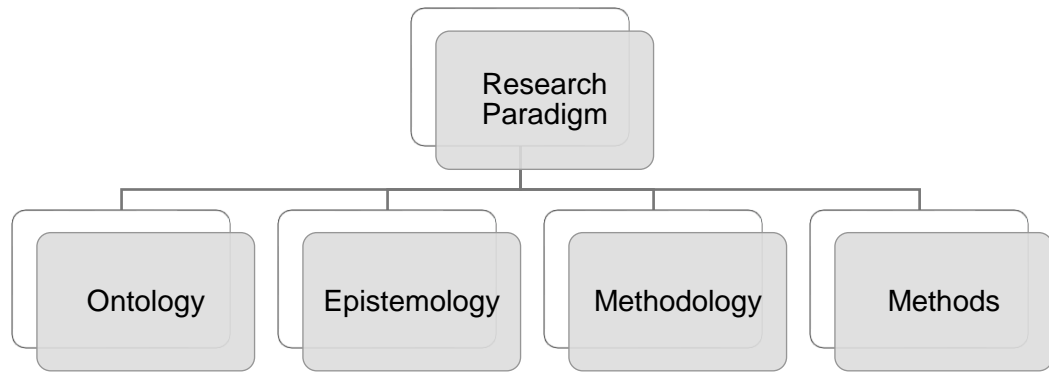


Figure 3.2. *Four Levels for Developing a Research Study*

According to Hall (2013), for example, positivism and post-positivism which is regarded as the successor of positivism are mainly associated with quantitative research and they argue that researchers can study the social world as they would study the natural world. They believe scientific knowledge exists independent of human existence and their main aim is to discover the single reality and create laws that are generalisable and govern the universe (Blaikie & Priest, 2017). It is required that researchers sustain full objectivity at all times and they conduct their inquiry without engaging in any non-scientific ways of gaining knowledge such as beliefs, feelings and thoughts (Crotty, 1998). However, such kind of approach falls short when it is applied to social sciences, a field that is concerned with society and the relationships among individuals within a society. Human autonomy means that the actions and beliefs of individuals are not easily predicted or pre-determined in the same way that scientific concepts follow specific laws. Although human behaviours and experiences are not always observable, they are still important and need to be considered (Mertens, 2014). On the other edge, constructivism, which is mostly associated with qualitative research, claims that reality is socially constructed by people and it bases its theory on understanding the world with the help of people living in and experiencing it (Mertens, 2014). As Crotty puts forward (1998), constructivism is relativist, not realist as opposed to positivism. That is, it does not represent a single reality as such, but it reports what people experience with respect to the phenomenon in question in a particular society. Truth is subjective, and it changes from person to person. Therefore, the knowledge and understandings need to be treated cautiously (Crotty, 1998). It is apparent that these two competing paradigms have distinct philosophical assumptions concerning research which are not compatible with MMR. As MMR researchers strive for combining both objective and subjective aspects of reality which encourage the use of quantitative and qualitative approaches, the above paradigms do not fully meet their expectations.

In response to the problem of paradigm incommensurability, it is suggested that there are three main stances that MMR researchers can adopt: a-paradigmatic stance, multiple paradigm stance and single paradigm stance (Hall, 2013; Teddlie & Tashakkori, 2009). Firstly, the a-paradigmatic stance claims that methods and paradigms are independent of each other and it focuses on convenience rather than engaging in discussions about philosophical underpinnings. Second, the multiple paradigm stance which Teddlie and Tashakkori (2003) divide into three categories (i.e., complementary strengths stance, the dialectical stance and the multiple paradigms stance) argues that researchers can choose more than one paradigm in their research. Lastly, the third and most popular paradigmatic stance among theorists is the single paradigm stance which is also referred to as an alternative paradigm stance (Hesse-Biber & Johnson, 2015). As mentioned earlier, some paradigms are associated with particular research methods (e.g., constructivism with qualitative research and positivism with quantitative research). Since there was not a particular paradigm associated with MMR until recently, this third paradigmatic stance originally aimed to come up with a paradigm which would inform MMR researchers (Tashakkori & Teddlie, 2010). However, this was later reframed by Greene (2007) as alternative paradigm stance which is conducive to or even requires mixing methods and does not incorporate any issues pertaining to incommensurable philosophical assumptions. According to Tashakkori and Teddlie (2010), the possible alternative paradigms for MMR researchers are pragmatism, transformative emancipation and critical realism. The current research is informed by *Critical Realism* on the grounds that it does not have the limitations that other paradigms possess and that it has the potential to provide a strong basis for mixed methods research (Creswell & Clark, 2017; Hall, 2013). The limitations of each paradigm and how critical realism deal with these limitations are discussed in detail below.

The transformative paradigm, suggested by Mertens (2003; Mertens, 2007; Mertens, 2010), is concerned with issues such as empowerment, inequality or oppression (Creswell & Clark, 2017). Although there is not an extensive amount of literature framing this paradigm, it is believed that it appeals to groups of researchers that include, but are not limited to, Marxists, feminists, members of lesbian, gay, bisexual and transsexual communities (Creswell & Creswell, 2018). It is emancipatory oriented and aims for a better change for those who are marginalized (Creswell & Clark, 2017). Considering the current study is not concerned with the issues of power and emancipation, it has been decided that transformative paradigm does not correspond well with the aims of the current study.

As for pragmatism, it has long been associated with MMR by most researchers in social sciences (Creswell & Clark, 2017; Johnson & Onwuegbuzie, 2004; Johnson, Onwuegbuzie, & Turner, 2007; Morgan, 2007; Morgan, 2014; Yvonne Feilzer, 2010). The doctrine of pragmatism is based on Dewey's 'philosophy of free choice' approach (Dewey 1933/2013). Its main emphasis is on research questions, and it is concerned with the relationship between actions and consequences (Biesta, 2010). According to Teddlie and Johnson (2009), pragmatism has gained increasing attention from MMR researchers because it rejects the either-or dichotomy led by the constructivism versus positivism debate. It offers a third choice which enables researchers to do 'what works' including merging quantitative and qualitative approaches together to address the real-world problems (Creswell & Clark, 2017). Therefore, it is generally regarded as "a leading contender for the philosophical champion of the mixed methods arena" (Greene, 2008, p.8). However, it should be noted that pragmatism is not in the same category as the other paradigms. As Biesta (2010) suggests, it is not a philosophical paradigm as such, but rather a "a set of philosophical tools that can be used to address problems" (p.97).

Although it is widely embraced by MMR researchers in social sciences and it offers a lot to researchers, there are two main limitations of pragmatism that still need to be addressed. First, it is criticised for its epistemological relativism (Cameron, 2011). Pragmatism gives primacy to practicality and underestimates the value of philosophical assumptions that guide research (Hesse-Biber & Johnson, 2015; Maxwell & Mittapalli, 2010). Given that ontological and epistemological assumptions provide information about how researchers understand the world and reality and how they think knowledge and reality can be defined and gained, this is an important weakness which needs to be considered. Second, it is based on a short-sighted practicalism (Cameron, 2011). A pragmatist researcher's starting point is a research question. Under the "dictatorship of the research question" (Tashakkori & Teddlie, 1998), the pragmatist embraces all research paradigms and research methods and use them on an ad-hoc basis to find an answer to the research question, which posits a serious threat to validity of research. According to Hesse-Biber and Johnson (2015), approaching research paradigms and methods in the same way is problematic because each paradigm or method consists of distinct features which need to be acknowledged and clarified in research. Also, some researchers are concerned about the extent to which predefined questions are evaluated critically and for whom the methodological choices are practical (Hesse-Biber & Johnson, 2015).

These limitations have steered the direction of some MMR researchers towards another paradigm called critical realism which is also known as scientific or transcendental realism (Blaikie & Priest, 2017). This paradigm has developed from the works of the contemporary philosophers Roy Bhaskar and Rom Harré (Danermark, Ekstrom, & Jakobsen, 2005). Although it has not received much attention in MMR until recently, it is believed to have the potential to resolve the aforementioned issues that the other paradigms have in MMR (Creswell & Clark, 2017; Maxwell & Mittapalli, 2010). Unlike pragmatism, it does not sidestep the paradigm issue by arguing that research methodology is independent of ontological and epistemological discussions, and it also acknowledges the value of both quantitative and qualitative approaches (Maxwell & Mittapalli, 2010). Critical realism proposes that it can combine and reconcile ontological realism, epistemological relativism and judgemental rationalism all of which are believed to be embedded to each other and mutually compatible (Bhaskar, 2007). In the next section, these ontological and epistemological assumptions on which the current research is based are outlined in detail.

3.1.1. Ontological and Epistemological Assumptions

Ontology is concerned with assumptions made about the nature of reality, being and truth (Teddlie & Tashakkori, 2009). It seeks answers to the questions such as 'What is reality?' and 'What can be said to really exist?'. As the most important manifestation of realism (Maxwell & Mittapalli, 2010), critical realism adopts the depth realist ontology which assumes that there are three levels of reality: the real, the actual and the empirical (Blaikie & Priest, 2017; Collier, 1994) (Table 3.1).

Table 3.1. The Three Ontological Domains of Critical Realism

	Domain of Empirical	Domain of Actual	Domain of Real
<i>Mechanisms</i>			√
<i>Events</i>		√	√
<i>Experiences</i>	√	√	√

Adapted from Collier (1994)

The empirical domain refers to direct and indirect experiences which can be observed or sensed by human beings. At the level of the actual lie both experiences and events that

are caused by overarching mechanisms. However, these mechanisms do not belong to this level as they are not actually realized (Collier, 1994). Lastly, the real is concerned with mechanisms that are independent of mind and society. Since mechanisms are real, but are not perceived, they can only be confirmed by causality (Collier, 1994).

The way critical realism identifies causality is different from that of positivism (Blaikie & Priest, 2017). Positivism adopts the successionist view which claims that it is sequences of events that can be observed, not the causes, so all that science needs to is to deduce causality from regularities between events (Blaikie & Priest, 2017). Critical realism, however, favours the generative view which proposes that investigating regularities between events is the essential part of the scientific inquiry, but it is also needed to explore mechanisms that link events to each other (Blaikie & Priest, 2017). Additionally, in social sciences, another term needs to enter the equation: the context (Figure 3.3).

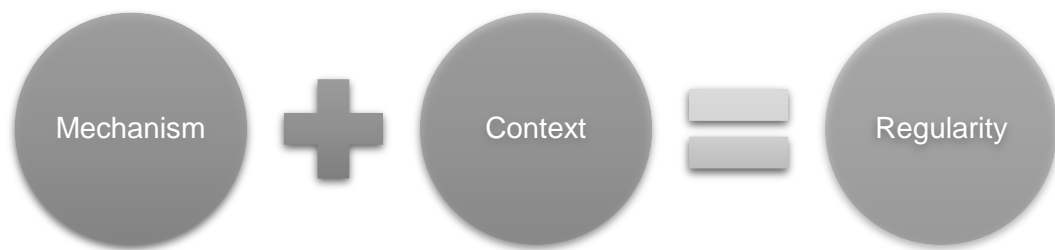


Figure 3.3. *Causality in Critical Realism - Adapted from Blaikie and Priest (2017)*

According to Blaikie and Priest (2017), in social sciences research, regularities between events, or relationship between concepts, are not independent from the contexts in which they occur. The reality that is identified in these particular contexts is under the influence of time and space. Therefore, it is important to specify the factors such as characteristics of the individuals and the cultural environment they live in. In the current study, the chosen context was the Turkish community and it was acknowledged at all phases of the research.

Following these ontological assumptions, it is important to clarify the underlying epistemological assumptions as well. Epistemology deals with issues concerning the theory of knowledge and its justification and the ways of knowing and learning about social reality (Teddle & Tashakkori, 2009). The questions that need to be answered are 'What is the nature of the relationship between the knower (the inquirer) and the known (or knowable)' and 'How can the reality be known?'. Critical realism adopts epistemological relativism and judgemental rationality (Archer, Bhaskar, Collier, Lawson,

& Norrie, 2007; Maxwell, 2012). It acknowledges that knowledge cannot be universal and human beings can only acquire the knowledge which is socially constructed and bound to change over time and across cultures. Thus, what can be learned about the world could just be an approximation to the truth. However, it is important to note that epistemological relativism does not bring about judgemental relativism in critical realism. According to Bhaskar (2007), critical realism features judgemental rationality which suggests that it is possible to judge between differences and form a reason. That is, there must be rational explanations counting for competing knowledge claims and it is researchers' responsibility to work on these.

In social research, the above process is facilitated by four forms of reasoning which are aligned to particular paradigms: induction, deduction, retroduction and abduction (Danermark et al., 2005). In critical realism, retroductive and abductive reasoning is contrasted to deductive and inductive reasoning. Deductive reasoning which is favoured by positivism requires developing hypotheses prior to research and aims to prove or disprove these. The limitation of this strategy in terms of critical realism is that it is based on pure logic for inference and it does not say anything new about reality beyond what is available (Danermark et al., 2005). In other words, it does not inform researchers about abstract structures and mechanisms that might contribute to understanding of a particular phenomenon. Unlike deductive reasoning, inductive reasoning, mostly associated with constructivism, assumes that knowledge can be gained making inferences from observations. Its main aim is to examine regularities and mechanisms in the observed events. Based on the generalisations made, researchers develop hypotheses that can be tested deductively. Inductive reasoning is also limited because knowledge can only be gained through empirical generalisations and regularities. It is neither empirically nor analytically certain (Danermark et al., 2005).

In critical realism, the key process by which aforementioned rational explanations are produced is either retroductive or abductive (Teddle & Tashakkori, 2009). Although abductive and retroductive reasoning is used interchangeably in some studies, researchers who take the critical realist path tend to differentiate between them. Therefore, these are treated as different entities in the current study as well. Retroductive strategy is a bottom-up approach which aims to uncover underlying mechanisms to reason observed regularities backwards (Blaikie, 2010). That is, the movement of inquiries is from a known regularity to an unknown explanatory mechanism (Blaikie & Priest, 2017). Research starts with an established theoretical or conceptual framework based on the context and possible mechanisms and it requires researchers to explain

the extent to which these provide an explanation using a combination of inductive and deductive reasoning (Blaikie, 2010). According to Blaikie (2010), the biggest challenge for this strategy is how to establish which mechanisms might provide the best answer as it is not based on formal logic. It is agreed that this can be achieved with disciplined scientific thinking aided by a combination of creativity, imagination, intuition and prediction (Blaikie, 2010).

The current study adopts abductive reasoning as a logic of inquiry. Like retroduction, abductive reasoning requires researchers to commence their inquiry with a theory in mind. However, abduction is not entirely based on guesswork as in retroductive strategy. It utilises formal logic as well (Danermark et al., 2005). According to Danermark et al. (2005), it is rare that social sciences inquiries conclude with something that has not been discovered before. Researchers generally aim to give a new meaning to already known phenomenon. They shed light on events and mechanisms that cannot be directly observed so that phenomenon is understood and explained in a new way. This kind of reasoning is known as abductive reasoning which can also be described as redescription or recontextualisation (Danermark et al., 2005). Abductive reasoning involves discovering the lay concepts, meanings and motives that are used by social actors and direct their behaviour (Blaikie, 2010; Blaikie & Priest, 2017). The lay descriptions and explanations are produced in an iterative way using both inductive and deductive logic. These are later informed by participants' perspectives through which the world in a particular context is described and understood. Based on these, researchers derive a scientific explanation (Danermark et al., 2005). The theoretical and conceptual frameworks that are proposed in the current study have been examined in detail in the previous chapter. The below section seeks to identify the methodology and the methods employed to address the research questions informed by the theoretical and conceptual frameworks.

3.2. Mixed Methods Research and Data Collection Methods

3.2.1. Definition of Mixed Methods Research

Before proceeding to outline the practical and scholarly reasons for choosing the Mixed Methods Research (MMR) design as the methodology of this PhD research, it is necessary to establish the foundations of MMR by providing a definition and exploring its nature and general characteristics.

MMR has emerged as a response to the incompatibility issue of quantitative and qualitative approaches and has gained increasing significance since its introduction as

'the third methodological movement' (Fraenkel, Wallen, & Hyun, 2012; Teddlie & Tashakkori, 2003; Teddlie & Tashakkori, 2009). It has been given many names such as integrative research, triangulated studies, multimethod research, blended research, multiple methods and mixed research (Johnson et al., 2007). However, *mixed methods research* has been the most commonly used term describing this movement (Johnson et al., 2007). Therefore, to be consistent with the majority of the literature, I have used MMR throughout this thesis. It needs to be noted that the word 'methods' in this term has a broader meaning as it actually refers to 'methodology' and includes the issues and strategies concerning the key elements of methodology (Johnson et al., 2007).

As Gray (2014) mentions, defining MMR is not a simple task since it is a relatively new concept and there are still some arguments and controversies concerning the essence of it. While some researchers provide a definition focusing on combining methods only, some view it as a wider concept and include other elements such as methodology and philosophy in their definitions (Creswell & Clark, 2017). In an attempt to put forward a comprehensive definition, Johnson et al. (2007) analysed 19 definitions provided by 21 leading methodologists in the field of MMR and suggested the following definition:

“Mixed methods research is the type of research in which a researcher or team of researchers combines elements of qualitative and quantitative research approaches (e.g., use of qualitative and quantitative viewpoints, data collection, analysis, inference techniques) for the broad purposes of breadth and depth of understanding and corroboration.”

Creswell and Clark (2017) further elaborate on this definition and add that MMR procedures are not designed randomly, but within a theory or philosophy. It is clear from these that MMR is more than simply using both qualitative and quantitative methods in the same study. It is a distinct methodological orientation with its own viewpoints, language and specific research techniques (Creswell, 2014; Creswell & Clark, 2017; Teddlie & Tashakkori, 2003).

According to Teddlie and Tashakkori (2010), one of the general characteristics of MMR is its methodological eclecticism meaning that researchers have the freedom to combine both qualitative and quantitative methods which they think are the most suitable tools for addressing their research questions or hypotheses. This key feature, however, has raised concerns among some researchers (Teddlie & Tashakkori, 2012). Teddlie and Tashakkori (2012) indicate that there are at least three issues surrounding the methodological eclecticism. First, critics question whether MMR researchers are

competent enough to conduct both types of research (Creswell & Clark, 2017; Teddlie & Tashakkori, 2012). Denzin (2008), for example, argues that “methodological bilingualism” might result in underestimating the distinct features of qualitative and quantitative research. Consequently, rigorousness of research methodology becomes questionable (Denzin, 2008). It is suggested that this weakness can be remedied to some extent by combining coursework and field experiences (Teddlie & Tashakkori, 2012), and there is already a growing literature concerning this (e.g., Christ, 2010; Earley, 2007). The second criticism concerns MMR researchers’ lack of imagination (Teddlie & Tashakkori, 2012). It is believed that they generally tend to choose questionnaires for the quantitative strand and interviews for the qualitative one and combine these in their studies. This could be seen essential during the first stages of MMR, but it is now needed that researchers become more innovative and introduce new complex combinations of research tools (Teddlie & Tashakkori, 2012). Recent studies have shown that researchers have been more open to integrating more complex and advanced techniques from the quantitative and qualitative approaches (for some examples, see Creswell & Clark, 2017). Therefore, this criticism is becoming unfounded. The last criticism is related to the paradigm issue (Teddlie & Tashakkori, 2012). The advocates of the incompatibility thesis believe that MMR researchers fail to seriously engage in philosophical discussions (Teddlie & Tashakkori, 2012). As already discussed in the previous sections, MMR researchers do not abandon the idea of research philosophy, but rather reframe the paradigm issue and reflect on how this can be addressed, specifically in MMR. Despite the above criticisms against MMR, the majority of the literature supports the use of MMR in social sciences for a number of reasons which are examined in the next section.

3.2.2. Rationale for Mixed Methods Research

It is one of the concerns in the field of MMR that researchers do not explicitly state their reasons for combining quantitative and qualitative approaches in their studies (Bryman, 2006; Bryman, 2007; Bryman, 2008). Bryman (2006), after having reviewed 232 social science articles, concluded that no rationale was given in twenty-five percent of the studies submitted to the content analysis. According to Bryman (2008), it is generally assumed that MMR is superior per se to single-method research. However, it is erroneous to believe the rationale for MMR is self-evident. Researchers are expected to clearly justify why MMR approach is suitable for their studies so as to sustain the quality of MMR studies (Bryman, 2008)

Based on Greene et al. (1989), Gray (2014) proposes that there are five main reasons to adopt MMR in studies: triangulation, complementarity, development, initiation and expansion. Triangulation means using multiple data sources in research to test the validity of findings. Using a single approach might be inconclusive because each approach has its own strengths and weaknesses. Therefore, researchers who want to provide full explanation can combine both approaches. This improves the quality of research as it offers a more reliable understanding of the same phenomenon. In a complementarity MMR study, qualitative and quantitative methods are utilised to measure different but related elements of the same phenomenon. It seeks clarification and interpretation of the results from one method with the results from the other. The third reason, development, concerns using the results of one approach to inform the development of the second approach. This enables researchers to obtain further information using a different method. Initiation helps explain the discrepancy in data. It focuses on the generation of new pathways in a research study which might require researchers to reframe their research questions when they discover contradiction in data. Lastly, expansion concerns broadening and widening the breadth and range of an inquiry. It provides a more complete picture of the same phenomenon and enhances the richness of findings.

The rationale for adopting MMR approach in this programme of research is that of triangulation, complementarity and expansion. The quantitative approach which was featured as questionnaires and experiments in the current research first aimed to examine whether there is a relationship between learners' gender stereotyped beliefs about foreign language learning and their L2 self-efficacy, L2 anxiety, and performance. This approach helped explore relatively stable beliefs and self-perceptions. Secondly, it aimed to experimentally investigate the impact of stereotype threat pertaining to learning another language on L2 performance. Thanks to the experiments, it was possible to examine learners' responses to different stereotype threat conditions. Semi-structured interviews which were chosen for the qualitative approach helped further explore the extent to which teachers, as an agent of socialisation, play a role in sustaining or legitimising these gender stereotyped beliefs. Semi-structured interviews provided the opportunity to dig into the impact of the context. Given the wide and varied focus of this study, MMR approach seemed ideal with the research questions and what the current research aimed to achieve.

3.2.3. Mixed Methods Research Design

It is also important for researchers to provide a transparent account of the MMR design, which is also known as typology, and its appropriateness to their research questions as this is one of the quality criteria for MMR (O'Cathain, Murphy, & Nicholl, 2008). The design of a MMR study is informed by the purpose for combining qualitative and quantitative approach because different purposes call for different types of designs. As noted by Greene (2008), there are three key dimensions which distinguish among the MMR designs. The first primary dimension concerns the interaction between the different methods (i.e., are the different methods being implemented interactively or independently). The second dimension refers to the status of the methods (i.e., is priority given to a particular method or do they have the same weights). Lastly, researchers need to decide on the timing of the different methods (i.e., are the different methods used concurrently or sequentially?). In order to help MMR researchers as to how to address these dimensions in accordance with their inquiries, a number of MMR designs have been suggested by different authors. Initially, it was reported that there were about forty MMR designs in the literature (Teddle & Tashakkori, 2003). As the field has expanded and developed over the years, these typologies have changed as well. According to Creswell and Clark (2017), the key ones among these are explanatory sequential design, the exploratory sequential design, the convergent parallel design.

In the explanatory sequential design, researchers begin their inquiry with the collection and analysis of quantitative data. This is followed by the qualitative strand which aims to build on the preceding stage (Creswell, 2014; Creswell & Creswell, 2017). This design allows researchers to expand the results gained from the quantitative strand or to explain nonsignificant, surprising and confusing results. Researchers who want to form groups for the qualitative phase can also use this design as it can facilitate their sampling strategy (Creswell & Creswell, 2017).

Regarding the exploratory sequential design, it has three phases. The first step, as opposed to the explanatory design, is to explore a problem through qualitative data collection and analysis. Following this, researchers can generate a quantitative measure, instrument or intervention (Creswell & Creswell, 2017). In the last stage, this new quantitative component is implemented. The convergent parallel design is different from the previous two designs in that qualitative and quantitative data is independent from each other and researchers collect both data concurrently (Creswell & Creswell, 2017). The idea is to bring all data together after the collection and analysis of each type of data. It enables researchers to have a complete picture of the problem and/or validate their findings (Creswell & Creswell, 2017).

In this study, the convergent parallel design which allows researchers to acquire different, but complementary data on the same phenomenon was adopted. As the nature of the design suggests, the quantitative and qualitative strands were implemented concurrently and with equal weight (Figure 3.4).

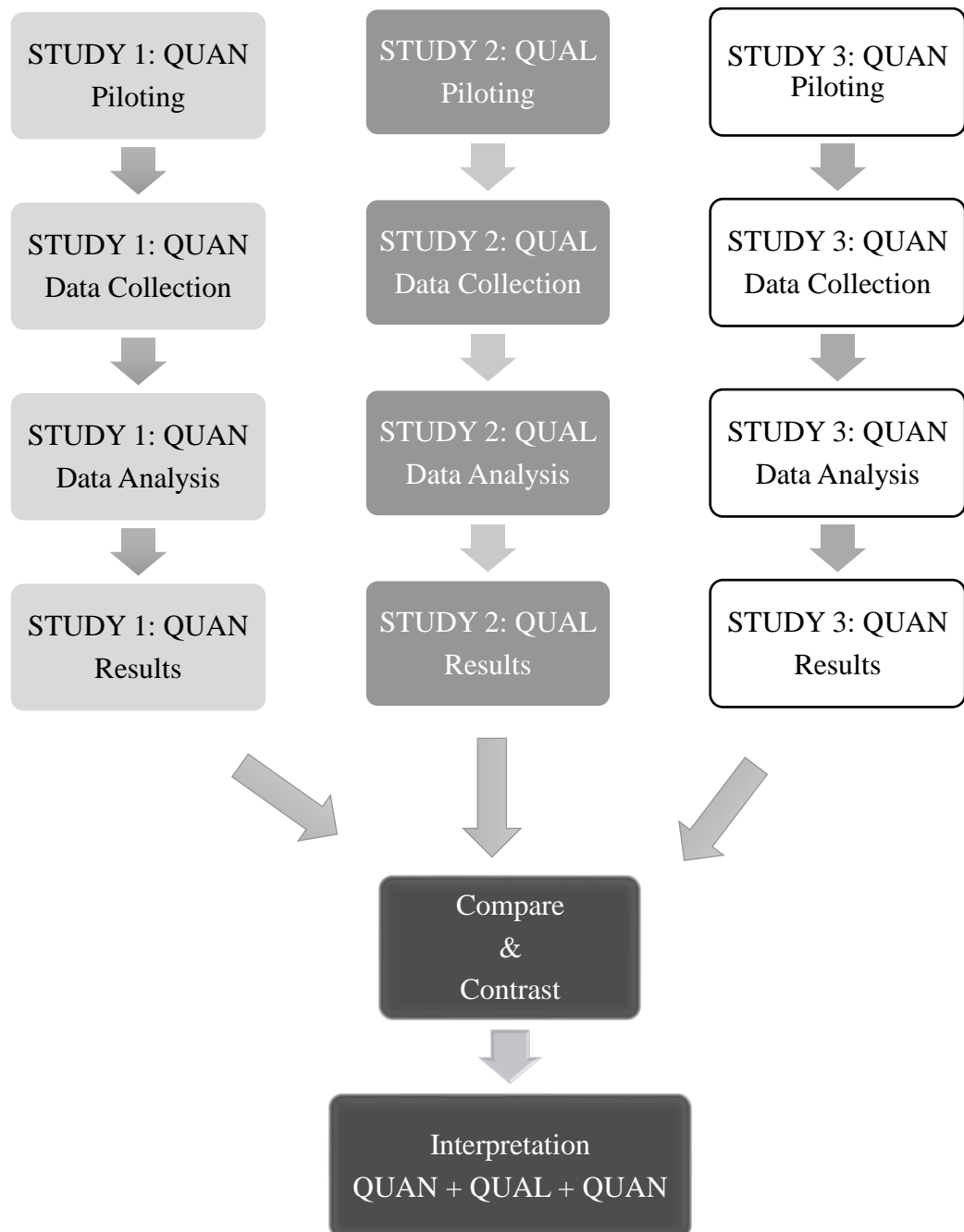


Figure 3.4. *The Convergent Parallel Design of the Current Study*

The strands were independent from each other during the collection and analysis stages and they were all mixed during the overall interpretation of the results as suggested by

Creswell and Clark (2017). Having discussed the overall design of the current study, the next section moves on to explain the data collection and analysis phases of each study in detail.

3.3. Setting and Participants

3.3.1. The Context of Turkey

Before proceeding to examine the situation of teaching and learning English as a foreign language (EFL) at university level in Turkey, which is the particular focus in the current study, it is necessary to shed light on the reasons for the importance of EFL in this country. There are two main reasons for EFL to be a significant concept in Turkey's agenda. The first reason concerns the country's socio-political situation. It occupies a strategic location lying between two major continents, Europe and Asia and having the vital crossroads between these continents. Also, it is a member of the North Atlantic Treaty Organisation (NATO) and an associate member to the EU. These indicate that Turkey has a significant position in the international arena. Due to its strategic and geopolitical status, more attention is needed to be given to teaching and learning English which is regarded as the lingua franca of the world and used as the main language for international communication in many areas such as science, technology and business. The second reason is that of globalization which has increased the power and dominance of English around the world. According to Tsui and Tollefson (2007), with the advent of globalization, two mediation tools have become crucial: technology and English, so all countries, regardless of their location and socio-political situation, strive to be equipped with adequate knowledge of these two skills. As being one of these countries, Turkey has been struggling to achieve the global standards of English for years now.

Although there have been various discussions concerning how to improve the quality of teaching and learning EFL in Turkey, there are still thorny issues that could not be solved (West, Güven, & Ergenekon, 2015). According to West et al. (2015), the overarching problem causing these issues is the inadequacy of English education in primary and secondary schools. Even after more than 1000 hours of teaching English (which is the sum of all the English classes that the students take until the end of Grade 12, the last year of High School), students still lack a good level English (West et al., 2015). Given the importance of English to Turkey, its underperforming in the area of English language teaching and learning in schools could be a threat to its economic and political development in the future (Vale, Ozen, & Alpaslan, 2013).

The deficiency in English in primary and secondary schools affects the quality of the Higher Education (HE) institutions as well (West et al., 2015). Students who do not have a good basis of English in primary and secondary schools find it challenging to learn it during their further studies because English Language Teaching (ELT) is more intensive and comprehensive at university level. Unlike the English classes in primary and secondary schools which are generally concerned with general English, ELT at university level is a broad concept which might include several main branches depending on the specific language programme. Among these come English for General Purposes and English for Specific Purposes (English for Academic Purposes and English for Occupational Purposes). In Turkey, there are more than 175 universities (state and private) and most of these have English preparatory programmes which offer such courses.

English preparatory programmes which last one full academic year (8-9 months) are not compulsory for all the students studying at a university. The necessity of them depends on the undergraduate programme in which the students are registered. The undergraduate programmes differ from each other based on their medium of instruction. While some use English as a medium of instruction (EMI) or Turkish as a medium of instruction (TMI), some others adopt a combination of these two (T-EMI). The English preparatory programmes are generally compulsory for those who pass the university entrance exam and are accepted at an undergraduate programme using EMI or T-EMI. Before these students start their undergraduate studies at university, they have to attend a preparatory course and pass a proficiency test at the end of the course.

As Turkey is a member of the Council of Europe, most of the English preparatory schools within Turkey adopt the Common European Framework of Reference for Languages (CEFR), a Council of Europe document published in 2001, and design the programmes in line with this document. The CEFR (2001) is an international standard for describing language ability. It aims to create a common basis for anything related to language learning such as language syllabuses, curriculum guidelines, examinations, and textbooks across Europe. It explains what language learners need to know in terms of knowledge and skills in order to use the new language effectively (CEFR, 2001). According to this comprehensive framework, there are six broad levels that language learners can achieve. These are illustrated below in Figure 3.5 (CEFR, 2001):

CEFR (2001) also provides a detailed global scale for each level (Table 3.2). According to the CEFR (2001), "Such a simple 'global' representation will make it easier to

communicate the system to non-specialist users and will also provide teachers and curriculum planners with orientation points.” (p.24).

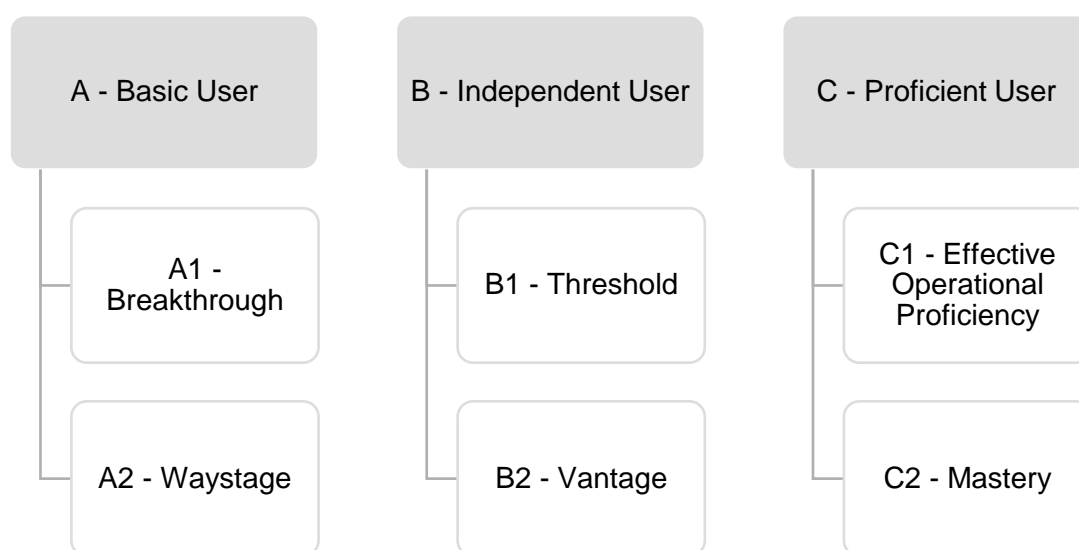


Figure 3.5. *Six Proficiency Levels in the CEFR*

In Turkish universities, most students start studying at university preparatory programmes with a CEFR level of A1+ and are expected to reach B2 level in a period of nine months. West et al. (2015) list three challenges that accompany such an approach. First, the threshold for entry for EMI students is very low and it is challenging for the students to reach the target level within nine months as they are expected to achieve a lot in a very limited time. Second, although most students are aware of the benefits of learning English in the long term, they are not motivated to learn it in a very intensive programme which needs them to attend 30-35 hours of English every week. That is, short-term difficulties outweigh the long-term advantages, which causes students' lack of motivation. Lastly, the classes in English preparatory programmes are usually compulsory, which makes students feel that they attend the classes because they have to, rather than because they want to. The discussions concerning how to address these issues and improve the quality of ELT at university level are still ongoing in Turkey.

Table 3.2. Common Reference Levels: Global Scale

Proficient user	C2	Learners at C2 level can understand everything they hear and read without any difficulty. They can summarise information from various spoken or written sources. They have the ability to express themselves spontaneously, fluently and precisely.
	C1	Learners at C1 level can understand a wide range or challenging and longer texts. They can recognise implicit meaning in the texts. They can express themselves without an obvious effort in searching for expressions. They are effective users of language in social, academic and professional environments.
Independent user	B2	Learners at B2 level can understand the main points of a complex text. They can interact with the other person with a degree of fluency and spontaneity. They can produce a clear, detailed text on a wider range of subjects. They can discuss the advantages and disadvantages of a chosen topic.
	B1	Learners at B1 level can understand the main points of clear standard input on familiar matters. They can talk about experiences and events, dreams, hopes and ambitions and briefly give reasons and explanations for opinions and plans. They can write a simple connected text on topics which are familiar or of personal interest.
Basic User	A2	Learners at A2 level can understand information related to their immediate environment (e.g., very basic personal and family information, phrases to describe locations, shopping etc.). They can interact with the other person provided that required information is simple.
	A1	Learners at A1 level can understand and use familiar everyday expressions and very basic phrases. They can introduce themselves, ask and answer personal questions such as where they live, people they know etc. They can interact with the other person if he/she speaks slowly and clearly.

3.3.2. Sampling Strategy

The participants were selected through the multi-stage sampling strategy. This strategy is the extended version of cluster sampling which is one of the probability sampling techniques (Gray, 2014). According to Ruane (2005), cluster sampling can be used when it is impossible or impractical to reach the whole population that is targeted. It is

convenient and economical because it provides researchers with a practical solution to recruiting a representative sample of very large research populations (Coolican, 2014). It allows researchers to divide large populations into separate groups, namely clusters. Then, they can choose one or some of these clusters to conduct their research and everyone in the clusters are sampled. However, it is not always possible and convenient to recruit everyone in the chosen cluster. The multi-stage sampling enables researchers to add a second or more stages to their sampling method. After choosing the clusters, researchers can narrow down the clusters further and choose a sample from each of the selected clusters (Babbie, 2013).



Figure 3.6. *Description of Seven Geographical Regions of Turkey*

In the current study, the target population was adults or young adults learning English as a foreign language in the HE institutions in Turkey which is divided into seven different geographical regions (Stage 1) (Figure 3.6). As this was a large population to recruit, it was decided to choose a cluster which could potentially be representative of the whole population. Therefore, the clusters were chosen from the HE institutions in Istanbul which is by far the largest city in Turkey as there are about 15 million people currently living in it (Stage 2) (TurkStat, 2018).

In Istanbul, there are about 50 universities (private and state) that could be targeted for potential participants. From among these state and private universities, further clusters were selected for each study on a random basis (Stage 3) (Figure 3.7).

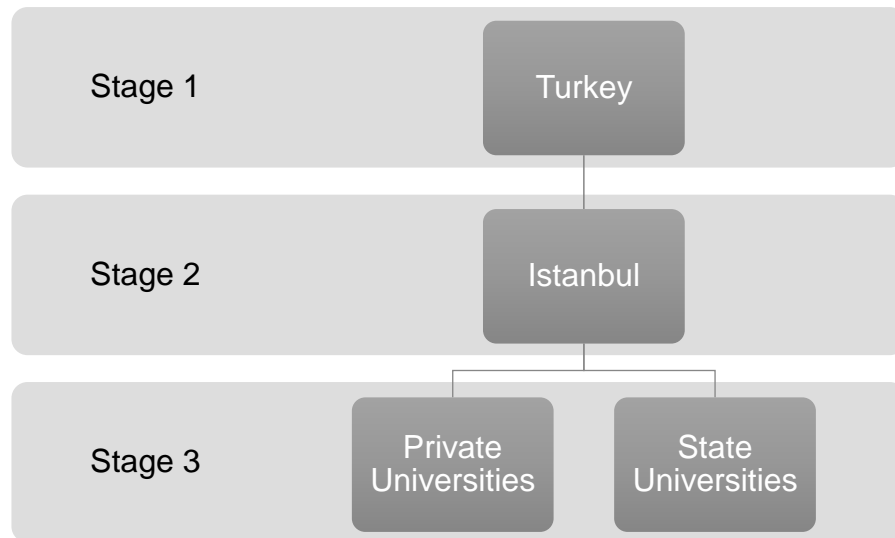


Figure 3.7. *Sampling Stages in the Current Study*

This kind of sampling might be criticised since clusters chosen may not be always a good representative of the whole population. Selecting the clusters from Istanbul addresses this problem to some extent because according to Turkish Statistical Institution (TurkStat), Istanbul has always received the highest amount of migration from all these regions of Turkey in 2014-2015 (See Figure 3.8). This means that there is a diverse population in Istanbul coming from the different parts of Turkey. Therefore, it would be possible to access a variety of different perspectives within the sampling frame.

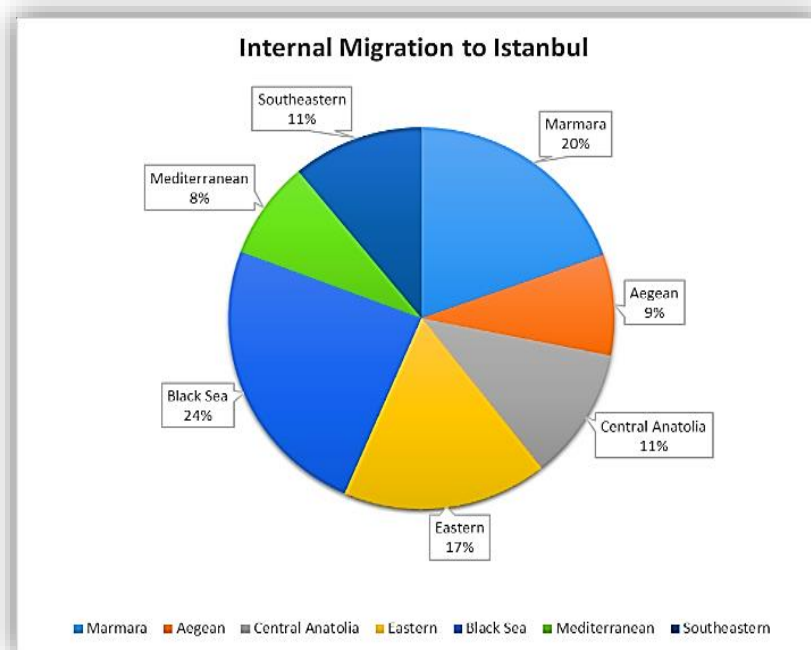


Figure 3.8. *Percentages of Internal Migration to Istanbul*

In this research, it was also necessary to conduct a pilot study, which will be described in Chapter 4. It is suggested that to improve the internal validity of the pilot studies, the participants should be chosen in exactly the same way as it is planned to be administered in the main study (Van Teijlingen & Hundley, 2001). Therefore, the participants for the pilot study were also chosen following the same stages as the main study.

3.3.3. Sample Size

The sample sizes for quantitative and qualitative approaches differ due to their distinctive natures. While a large sample size is needed for a quantitative approach, a relatively smaller sample size is acceptable in a qualitative approach (Gray, 2014). As each study in the current research was unique and needed to be treated as a separate entity, the sample sizes differed across the studies in the current research.

Piloting (All studies)

Although piloting is a good practice in social sciences, there is not a precise guidance in literature concerning the sample size in pilot studies (Johanson & Brooks, 2010). Nevertheless, there are some recommendations, although not exhaustive, that are intended to help researchers with their decisions. Based on congruence between rotated population and sample factor loadings, Comrey and Lee (1992) recommended that the sample size ideal to validate a scale should be based upon the following guidance: 100 participants = poor, 200 participants = fair, 300 participants = good, 500 participants = very good, ≥ 1000 participants = excellent. As the main purpose of piloting in this study was to validate the questionnaires which were developed to address this research's aims and questions, a total of 336 students (153 Female/183 Males) were recruited to test the validity and reliability of the data gained using the questionnaires as well as the feasibility of Study 1 as a whole.

In Study 2 and 3, it was decided to follow the suggestions of Connelly (2008) who states that in extant literature, a pilot study sample is generally 10% of the sample planned for the main study. Based on these recommendations, in the second pilot study which aimed to test the interview protocol and questions, two teachers (one male and one female) and two learners (one male and one female) were recruited because the intent was to interview 16 teachers and 16 learners in Study 2. In the last pilot study which was done for the experimental part, Study 3, there were three different groups (two experimental and one control groups). Therefore, three learners were recruited for each group.

Study 1

The sample size for this study was calculated using Creative Service Systems Sample Size Calculator. This programme is available at <http://surveysystem.com/sscalc.htm>. It requires researchers to input their desired confidence level, confidence interval and the population size, and the sample size is calculated accordingly. According to Cohen et al. (2018), these three variables are crucial in determining the correct sample size. As for the size of the targeted population in Turkey, there is not an exact number of the students attending the English preparatory classes, but it is possible to access the total number of the students enrolled in the undergraduate programmes. According to the Turkish Council of Higher Education Statistics (2018), 4,071,579 students (1,869,240 females, 2,202,339 males) were registered at an undergraduate study in the last academic year (2016-2017). Although not all of them were entitled to study English on a compulsory basis at preparatory classes, the figure of 4,071,579 was used as a starting point. The confidence level refers to a percentage (usually 95 per cent or 99 per cent) which indicates the extent to which researchers are sure about their findings. It is a consensus that a 95 per cent confidence level is conventionally regarded as appropriate (Cohen et al., 2018). Therefore, the sample size was calculated with 95 per cent for the confidence level and ± 5 per cent confidence interval which shows the degree of variation or variation range (e.g. ± 1 per cent, or ± 2 per cent, or ± 3 per cent). Based on this calculation, it was decided that approximately 400 students would be enough to be surveyed with three questionnaires in Study 1. This was in line with Gay et al. (2003) who suggested that when the population has about 5,000 or more people, sample size of 400 should be adequate. Following these suggestions, 701 students (male = 346, 49.6%; female = 355, 50.6%) were recruited in Study 1.

Study 2

Interview with the teachers

According to Seidman (2013), there are two criteria for deciding upon the sample size in interviewing: sufficiency and saturation of information. It needs to be ensured that the sample size is large enough to reflect the variety of the target population. In the current study, it was proposed that 16 teachers who teach to the students included in Study 1. To have good balance of gender, eight of these were planned to be males, and the rest would be females. It was also ensured that there were enough teachers representing each university selected for Study 1. This sample size was subject to change based on the second criteria, namely the saturation point. According to some scholars (e.g., Brinkmann & Kvale, 2015; Seidman, 2013), when there is no new information that researchers can get from the interviews, there is no need for them to continue interviews

further as this means that they are at the saturation point. In this study, the saturation point was reached after interviewing 16 participants as proposed.

Interview with the students

Similar to the interviews with the teachers, the saturation point was crucial in determining the sample size for the interviews with the students as well. As planned initially, 16 students (8 females and 8 males) were interviewed and this was sufficient to reach the saturation point. It is important to note that some scholars might prefer adopting a power analysis approach to decide upon the sample size in interviewing (e.g., Fugard & Potts, 2015). However, according to Fugard and Potts (2015), such an approach is limited in that it does not take the design of interviews (e.g., the length) into account. As such, the sample size determined by the power analysis would need to be adjusted considering the length of the interviews. Given that, the current research followed Seidman's (2013) approach which is widely accepted by qualitative researchers.

Study 3

The sample size of Study 3, the experimental part, was decided based on the previous studies which suggest that researchers should aim for a minimum sample size of 20 participants for each experimental and control group (Simmons, Nelson, & Simonsohn, 2011). Therefore, the aim was to recruit at least 60 students (20 students per group). The students had to be different from the ones who participated in Study 1 and Study 2 due to the nature of the experimental study. This is explained in detail below.

3.4. Data Collection

3.4.1. Study 1: Quantitative Approach (Closed-Ended Questionnaires)

In Study 1, quantitative data were collected using three closed-ended questionnaires which assessed learners' gender stereotyped beliefs about foreign language learning, self-efficacy and language anxiety respectively. The development and validation processes of each questionnaire will be covered thoroughly in Chapter 4.

Questionnaires are significant and one of the most commonly used data collection tools (Cohen et al., 2018; Gray, 2014). As Gray (2014) points out, the popularity of questionnaires is based on their advantages. The main advantages are as follows:

- Questionnaires are cost-effective in terms of time and money. They are easy to deliver to large groups of people at relatively low cost (Dörnyei & Taguchi, 2009; Gray, 2014).

- They are practical as they can be administered without presence of researchers (Wilson & McClean, 1994).
- They are versatile because they can easily be adapted to various people, situations and topics (Dörnyei & Taguchi, 2009).
- Analysing the data is comparatively quick and straightforward (Gray, 2014; Wilson & McClean, 1994).
- Unlike interviews, they are more reliable and consistent as they are free of researcher bias in the process of data collection (Dörnyei & Taguchi, 2009).

However, there are also some drawbacks associated with this type of data collection methods. According to Dörnyei and Taguchi (2009), some might think that data gained through questionnaires are simple and superficial. This happens especially when questionnaires are lengthy because people do not always want to spend a long time on completing them. Also, lengthy questionnaires might create a fatigue effect which means participants might fill in questionnaires inaccurately due to tiredness and boredom. Therefore, it is suggested that questionnaires should not be more than six pages (Gillham, 2007). As noted, the questionnaires in this study were five pages long in total and one of these pages included the demographic questions only. Even if the length of questionnaires is about right, there are still some limitations which cannot easily be controlled by researchers. For example, participants might misunderstand the information given or simply give inaccurate and misleading information on purpose (Dörnyei & Taguchi, 2009; Gray, 2014). As Gray (2014) states, researchers are not in the position to understand these kinds of responses or correct respondents' mistakes. However, it is possible to address this kind of limitations in interviews because researchers can observe participants' body language or detect any changes in their verbal tones (Gray, 2014). As mentioned earlier, one of the data collection methods in this study was interviews and they aimed to eliminate the aforementioned weaknesses pertaining to questionnaires.

According to Cohen et al. (2018), there are three types of questionnaires: structured, semi-structured and unstructured. The decision about which one to choose depends on the sample size of studies. If researchers aim for a large sample size, it is suggested that questionnaires should be more closed and numerical. However, if the sample size is smaller, a less structured questionnaire can be used (Cohen et al., 2018). The three types of questionnaires differ from each other in terms of the items found within them. As Gray (2014) mentions, a questionnaire can comprise either open or closed questions.

Open questions (i.e., How? What? Why? etc.) allow participants to provide their responses without any limitations of pre-set answers. Also, they are easy to set up and have the potential for richness of responses. However, they take a longer time to process and analyse. Especially with a large sample size, they become more challenging. In contrast, closed questions are more difficult to generate, but they are more straightforward to handle at the data processing and analysing stages. As they are more focused compared to open questions, the chance that researchers might end up with unexpected and surprising results that are difficult to compare across different groups is slim. In this study, the target was to reach about one thousand students including both the pilot and the main studies. Given the advantages of using a more structured questionnaire, it was decided that the questionnaire items in the aforementioned three questionnaires should be closed and numerical rather than open and word-based as found in less structured questionnaires. Closed questions can take many forms, namely listing questions, category questions and ranking questions. The questionnaires in this study were designed using ranking questions which are also known as scale or rating (Gray, 2014). As a common type of this category, Likert-type scales on which participants indicate how strongly agree or disagree with a series of statements were developed and implemented to measure learners' gender stereotyped beliefs, self-efficacy and anxiety.

3.4.2. Study 2: Qualitative Approach (Semi-Structured Interviews)

Questionnaires and interviews are often believed to be more effective when used together in a MMR study as they compensate each other's weaknesses (Kendall, 2008). While questionnaires allow researchers to investigate patterns among large groups, interviews help gather more in-depth information about participant attitudes, thoughts, and actions (Kendall, 2008). As Vygotsky (1986) noted, the words people use while presenting their stories are the microcosms of their consciousness. Researchers can utilise this consciousness to access to the most complicated social and educational issues because these issues are the results of the concrete experiences of people (Seidman, 2013). In this study, interviewing was implemented to explore foreign language learners' and teachers' perceptions and beliefs, which was not possible to gain through closed questionnaires.

Interviews can be described as *"an inter-change of views between two persons conversing about a theme of mutual interest"* (Brinkmann & Kvale, 2015, p.4). They are generally divided into three categories: structured interviews, semi-structured interviews and unstructured interviews (Brinkmann, 2014). Brinkmann (2014), however, highlights that there is no such thing as completely structured or completely unstructured interviews

since people might give more information than needed in a structured interview or researchers might be able to steer the direction of an unstructured interview as they know what the conversation needs to cover (Brinkmann, 2014). Therefore, the distinction between the three types of interviews needs to be treated as a continuum ranging from relatively structured to relatively unstructured interviews (Brinkmann, 2014). Kendall (2008) further differentiates between these types and call semi-structured and unstructured interviews 'qualitative interviews' as opposed to structured interviews.

Structured interviews are similar to questionnaires in that they require researchers to ask participants questions in a standardised way so that the answers can be compared across participants and explained in numbers (Brinkmann, 2014). In these interviews, there can be little or no variation in questions. This means that researchers have little freedom to change the prearranged content and procedures of the interviews (Cohen et al., 2018). As it only allows for limited participant responses and there is not scope for follow-up questions to responses that require further elaboration, they are of little use if depth is required. Unlike structured interviews, unstructured interviews provide researchers with a more open situation, flexibility and freedom (Cohen et al., 2018). They are mostly used when in-depth information is required, or little is known about the phenomenon in question (Gill, Stewart, Treasure, & Chadwick, 2008). However, unstructured interviews might be time-consuming, and the data collected is difficult to handle because in this type of interviews as there is little or no guidance for researchers (Gill et al., 2008). Researchers do not plan an interview schedule in advance. They start their inquiry with an opening question, a topic or a theme and they proceed based on the conversation they have with participants (Gill et al., 2008). Since each participant brings their own thoughts, feelings and experiences with them, interviews might differ from each other which affects the generalisability and representativeness of the data collected using unstructured interviews.

Semi-structured interviews, as the most commonly used type of interviews in social sciences, have the characteristics of both structured and unstructured interviews. Although researchers prepare their key interview questions in advance in semi-structured interviews, they are allowed to ask follow-up questions to pursue an idea or response for further detail (Brinkmann, 2014). In this type of interviews, there is more room for dialogue and other changes (Seidman, 2013). Hence, qualitative data was gathered using semi-structured interviews in the current study. Through these interviews, it was aimed to shed light on whether teachers as an agent of socialisation in a learning environment play a role in sustaining and legitimising gender stereotyped beliefs in

foreign language learning and the extent to which their beliefs are related to language learners' own gender stereotyped beliefs. To address this research question, both foreign language students and teachers were interviewed. To sustain the consistency across the interviews, the interview questions for both learners and teachers were categorised under three main headings: gender stereotyping, self-efficacy and language anxiety. Through interviewing teachers, I aimed to disclose the ways in which they might hold gender-biased beliefs in terms of language learning, what impact this might have on their practice in the classroom, potentially how these beliefs systems were formed and how they are communicated. The interviews with the students allowed me to discover these patterns more from a different perspective (see Chapter 5 for more details).

The efficiency of interviews depends on a number of factors one of which is the purpose of researchers (Seidman, 2013). As Seidman (2013) explains, interviewing is not appropriate if the main aim of researchers' inquiry is to measure participants' behaviours. In such cases, another method such as observations is believed to be more suitable (Kendall, 2008; Seidman, 2013). However, this does not mean that researchers cannot discuss participants' behaviours during interviews. Interviewing enables researchers to access participants' subjective understanding of their behaviours rather than an objective account about these behaviours (Kendall, 2008; Seidman, 2013). The second factor that affects the success of interviews concerns the relationship between the interviewee and interviewer (Kendall, 2008). According to Kendall (2008), it is difficult for researchers to establish rapport with participants and gain understanding as they are strangers who meet for the first time. To overcome this issue, it is expected that researchers' have a set of interpersonal skills (Kendall, 2008). However, thinking that the success of an interview depends on the personality of researchers imply that researchers are either good at interviewing or not (Seidman, 2013). It is suggested that the importance should also be placed on interviewing technique (Seidman, 2013). In line with this argument, Seidman (2013) provides some techniques that researchers need to use. Basically, interviewers need to listen more and talk less, follow up on what the participant says (e.g., asking for clarification, asking for more information), ask real questions (e.g., avoiding leading questions, asking open-ended questions), avoid interrupting the participants, limit their own interaction (e.g., sharing their own experiences only occasionally, avoiding agreeing or disagreeing with the participants' responses), use an interview guide cautiously and tolerate silence. All these suggestions were followed throughout the interviews which were conducted in Study 2 (see Chapter 4 for more information).

3.4.3. Study 3: Quantitative Approach (Experiments)

Differently from the other studies, Study 3 experimentally investigated stereotype threat with respect to foreign language learning. In this study, unlike Study 1 which focused on stereotype endorsement and its relation to students' self-efficacy, anxiety and L2 performance, the focus was on the effects of stereotype threat on these variables.

Experiments are regarded as one of the most effective ways of gathering evidence to support causal theories (Coolican, 2014). They allow researchers to eliminate the different explanations of observed relationships between independent and dependent variables and isolate cause and effect relationship that they aim to investigate (Coolican, 2009). Experiments consist of dependent and independent variables. An independent variable (IV) is the variable which researchers manipulate in their experiment. A dependent variable (DV) is the variable which changes depending on the manipulation of this independent variable. In this study, the IV was stereotype threat which was manipulated through different treatment conditions. The dependent variable was the learners' L2 performance which was assessed using a 12-minute English listening test. Learners' self-efficacy and anxiety were the mediators which were assessed using the same questionnaires used in Study 1. After completing the questionnaires, both control and experimental groups were asked to attend the English listening test. The test and assessment were based on the criteria suggested by the Cambridge English. The analyses were examined to investigate the extent to which different stereotype threat conditions affected the learners' self-efficacy beliefs, anxiety and performance.

According to Hole (2012), two of the commonly used experimental designs are 'post-test only/control group' design and 'pre-test/post-test control group' design. The design of this study was a post-test-only, equivalent control group design. As the nature of stereotype threat suggests, it is not necessary to know the participants' prior perceptions or performance. Also, having a pre-test might run the risk of affecting the way students completed the questionnaires and the final test. By including a pre-test, participants are sensitized to what is being studied which might induce demand characteristics and reduce the validity of measures. Therefore, for the purposes of this study, a pre-test was not conducted. As Hole (2012) indicates, in a post-test equivalent groups study design, there are experimental and control groups and participants are randomly allocated to each group. This ensures that experimental and control groups are equivalent. After each group receives different treatments, the results of the post-test are compared. In the current study, there were three groups: two experimental groups (i.e., self-as-target

threat and other-as-target threat conditions) and one control group. The control group did not receive any stereotype threat manipulation. The participants in the experimental groups were given two different instructions to assess the effects of stereotype threat manipulation. One of the instructions corresponded to *self-as-target threat*, whereas the other referred to *group-as-target threat*. The instructions were different from each other to see whether they affected dependent variables. These are covered at length in Chapter 6 (see Section 6.3).

Coolican (2014) notes that an effective experimental study addresses issues of validity and reliability. As for reliability, there are two factors that need to be considered. The first of these concerns sample size. Experiments using a small sample size suffer from a lack of power which refers to the ability to detect the existing experimental effects (Hole, 2012). According to Simmons et al. (2011), researchers should aim for a minimum sample size of 20 participants to avoid such limitation in their research because smaller samples are not powerful enough to determine most effects. In line with this suggestion, I aimed to recruit 20 male EFL learners for each group in the current research. As Study 3 consisted of three groups, it was aimed to recruit 60 male participants in total which is comparable to the previous stereotype threat studies (e.g., Nadler & Clark, 2011; Pennington, 2016).

The experimental design suggests that participants should not know what the real aim of the study is. Therefore, the participants in Study 3 were different from the ones who attended Study 1 and Study 2. All the participants were informed about the aims of the study after they attended the experiment. The second factor affecting reliability is the extent to which the phenomenon being investigated is stable (Hole, 2012). As Hole (2012) states, some phenomena are more stable than others. While some change depending upon the external factors such as the time of day and current mood of participants (e.g., state anxiety), some others are more stable and are not affected by such factors easily (e.g., trait anxiety). Therefore, researchers need to be wary when making generalisations based on the results gained through experiments.

The threats to validity can be categorised under three headings: time threats, group threats and participant reactivity threats (Hole, 2012). Time threats can cause problems for studies which take place over a prolonged period of time. For example, there might be specific events which occur between the first and second measurement and these could affect the results. Also, the results might be contaminated by maturational changes. If the study lasts a few months, participants' thoughts, feelings and

performance might improve regardless of treatment. Furthermore, if tests are implemented repeatedly, taking one test might influence the outcomes of the second test. Such threats were not considered in the current study as the experiments were not extended over time. The participants were asked to complete the questionnaires and the test at once. The only time threat that was relevant to the current study is related to instrument change. As Hole (2012) mentions, the way researchers conduct their experiment might change over time. While they might be less competent during the first experiments, they might become more careful and precise in the following experiments. The change in researchers' demeanour might result in changes participants' responses. It is suggested that this problem can be addressed by standardising the instructions given and conducting the experiments in different orders (e.g., conducting one experiment with the control group which would be followed by an experiment with the experimental group). As suggested, in this study, the order of the experiments was randomized. Also, as the instructions for each group were prepared before the experiments started, all groups received the same standardised instructions regardless of the order the experiments conducted.

Group threats relate to strategies for selecting participants (Haslam & McGarty, 2014; Hole, 2012). Ideally, the participants chosen for each group should be equal. However, participants within the groups differ from each other in many respects as they all have their own characteristics and experiences. In this study, all participants were chosen from among the learners studying English as a foreign language at different universities in Istanbul. Although they were equal in this respect, there were some other differences between the learners such as their background (e.g., not all students were from Istanbul), their English level (e.g., even in the same class, there were some very successful students and there were also learners who were lower than the rest of the class) and their motivation to participate in this study (e.g., although some were eager to contribute to the study, some agreed to participate only to escape the class). It is believed that such differences do not contaminate the results on condition that participants are randomly allocated to experimental and control groups. In line with this suggestion, the participants in Study 3 were recruited and allocated to the groups on a random basis. It is important to note that when random allocation is not used, researchers need to be careful while interpreting the results as it is difficult to be certain whether the differences in the results are because of the different instructions given or the group differences.

Lastly, researchers need to consider the participant reactivity threats as well. Participants might feel anxious about being evaluated by the experimenters which would affect the

way they complete the test (Hole, 2012). In the current study, participants were asked to complete the questionnaires and the test in their classrooms. Although there were informed that the results would only be seen by the researcher, they might still feel that they would be shared with their teachers. The experimenters themselves might influence the results as well. According to Hole (2012), researchers' behaviours might differ across the groups unintentionally as they might have certain expectations from each group. However, this is not regarded as a serious problem (Haslam & McGarty, 2014). Even though the participants are aware of the experimenter's expectations, they have the freedom to confirm and disconfirm these. Nevertheless, researchers are suggested to be careful not to disclose the whole story behind the experiments (Haslam & McGarty, 2014). In alignment with this suggestion, the aims of the study were not shared with the students until they had completed the English test and the questionnaires.

3.5. Data Analysis

As mentioned earlier, the results of the studies were first analysed separately, and the analyses were merged after the analysis of each study was completed (see Figure 3.3 above). Below, the data analysis procedures are set out in detail.

3.5.1. Close-Ended Questionnaires

The data from Study 1 was prepared for analysis using IBM Statistical Package for the Social Sciences (SPSS) (Version 24). All the data gained from the three questionnaires was entered into SPSS and it was checked for any data entry errors or any coding issues. All the missing data were identified and replaced with 999. The descriptive statistics such as means, ranges, standard deviations, variances, minimum and maximum variables, and measures of kurtosis and skewness were calculated using SPSS. The SPSS output file was then transferred to a *.dat* file which can be read by *Mplus* Version 7.4. (Muthén & Muthén, 1998-2017)

All the other analyses were done using *Mplus* 7.4 which is one of the ideal software packages for structural equation modelling (SEM). The reliability of the data gained from the three questionnaires was assessed using SPSS. As for the validity, a series of analyses were run on *Mplus* using maximum-likelihood estimation (Muthén & Muthén, 2013). The process of data analysis and results are discussed in detail in Chapter 4. Any missing data was handled using full information maximum likelihood (FIML) in *Mplus* 7.4 (Muthén & Muthén, 2012). According to Arbuckle (1996) and Enders and Bandalos (2001), FIML is more efficient to deal with missing data than the traditional methods such as listwise and pairwise deletion. FIML computes parameter estimates in the model on

the basis of all available data, including the incomplete cases. It has a number of theoretical and empirical advantages such as reducing bias resulting from missing data and preserving the overall power of the analysis (Graham, Van Horn, & Taylor, 2012).

3.5.2. Semi-Structured Interviews

Qualitative data gained through interviews was analysed using thematic analysis. In this method, qualitative data is examined in order to recognise and organise patterns based on their content and meaning (Willig, 2013). Braun and Clarke (2006) suggest that thematic analysis should be regarded as a foundational method for qualitative analysis. This is mostly because it is suitable for a wide range of materials such as transcripts of interviews, personal diaries, letters and blogs (Willig, 2013).

According to Braun and Clarke (2006), there are six main phases of thematic analysis: familiarising with the data, generating initial codes, searching for themes, reviewing themes, defining and naming themes and producing the final report. I familiarised myself with the data during the process of transcribing which was followed by the second step, generating the codes. Initial codes are determined either deductively or inductively (Willig, 2013). In a deductive thematic analysis, researchers have a pre-existing coding frame. In an inductive thematic analysis, however, the coding frame is not a predetermined one. It emerges from the available data and requires researchers to adopt a bottom-up approach (Willig, 2013). In this kind of approach, themes are grounded in the data and researchers' theoretical assumptions are not reflected in the analysis. In contrast, a deductive approach to thematic analysis allows researchers to use some form of template which is informed by the relevant literature. In the current study, both deductive and inductive approaches were adopted. As noted by Willig (2013), it is possible to use the combination of these two approaches. While researchers can have a template to use as a priori, they can accept any new themes that emerge in the analysis and merge both the initial and newly emerging themes to generate a comprehensive description of the data.

Following the coding process, the recurrent themes were identified. It is important to note that codes are different from themes (Willig, 2013). As Willig (2013) states, there might be more codes than themes in a piece of qualitative data. After determining the themes in the interviews, the themes were interpreted, and an explanatory framework based on the important themes were constructed. The final step in the analysis involved comparing the quantitative data with the qualitative one and finding the links, if there are any.

3.5.3. Experiments

Similar to Study 1, all the data were prepared for analysis in SPSS. The differences between the three groups, namely the experimental and control groups were analysed using analysis of variance (ANOVA) which is a statistical method used to test differences between two or more means (Haslam & McGarty, 2014).

As mentioned earlier, there were three groups in Study 3 and the main difference between the groups were the instructions given to each one. The control group was primed with *"In today's session, I want to assess various factors relating to language learning. This task is not diagnostic of your ability."* The experimental group was primed with *"In today's session I want to assess your language ability by asking you to fill in two questionnaires and attend an English listening test. Typically, females have been shown to outperform females on this specific test. Your performance on this will therefore be used to help us establish your personal foreign language ability."* This instruction aimed to prime self-as-target threat. The other experimental group was primed with *"In today's session, I want to assess language ability for males and females by asking you to fill in two questionnaires and attend an English listening test. Typically, females have been shown to outperform males on this specific test. Your performance on this will therefore be used to help us establish males' language ability."* This instruction aimed to prime group-as-target threat. Using the one-way analysis of variance (ANOVA), it was determined whether there were any statistically significant differences between the means of these three independent (unrelated) groups.

3.6. Data Integration

According to Creswell and Clark (2017), there are four aspects of integration that need to be addressed by researchers: the intent of integration, the data integration procedures, the representation of the integration results and the interpretation of integration results. The intent of integration in the convergent parallel design is to provide a comprehensive understanding of the data and validate and confirm the results. This can be done either by making a side-by-side comparison between the qualitative and quantitative data or by transforming one of the data sets into another data structure (e.g., transforming qualitative data into quantitative one). In the current study, the data were integrated using the former method. The comparison made by following the primary data analysis integration procedures outlined below:

- Each sets of data were analysed using the appropriate data analysis techniques (e.g., qualitative data was analysed using the qualitative data analysis techniques).

- Once the results were obtained from each analysis, they were analysed further to investigate the common concepts.
- Quantitative and qualitative results were compared for each concept.
- The results were examined to determine whether they confirm, disconfirm or expand each other.
- Any disconfirming results were subject to further inspection to find some kind of clarification.
- All the confirming, disconfirming and expanding results were interpreted to enhance the understanding of them.

The third step was to represent the integrated results. This could be done via a joint display or a narrative discussion (Creswell & Clark, 2017). In this study, the integrated results were presented in a narrative discussion (see Chapter 7 for General Discussion). The order of the discussion was aligned with the order of the studies. That is, the results from Study 1 was followed by Study 2 and 3. Lastly, after presenting the integrated results, these were further examined and interpreted to answer the research questions.

3.7. Ethical Considerations

This PhD project adhered to the Ethical Guidelines for Educational Research by British Educational Research Association (BERA, 2011) and Edge Hill's Framework for Research Ethics (FRE) (2016). As the project included such instruments as questionnaires and interviews alongside an experiment all of which required the participants' permission, the project was subject to ethical approval by the Faculty of Education Research Ethics Committee and it was obtained prior to the commencement of data collection (see Appendix 3.1. for the ethical approval letter). Adherence to BERA (2011) and FRE meant that the following responsibilities were considered throughout this study:

3.7.1. Voluntary Informed Consent

BERA (2011) clearly explains that even if the research is undertaken somewhere outside of the UK which was the case in the current research, researchers are liable to the same guidelines. One of the prerequisites for researchers is to prepare a voluntary informed consent which outlines the aims and objectives of their inquiries. Based on BERA's (2011) guidelines, a voluntary consent form along with a participant information sheet which gave information about the study was prepared and used to seek consent from the local authorities in Turkey (i.e., the head or director of the schools of foreign

languages based in Istanbul) (see Appendix 3.2 and 3.3). Although the authorities gave their consent, it was still needed to get the consent of the individual participants as they were the ones who would be interviewed, asked to complete the questionnaires or participate in the experiments. The participants were asked to read the document and agreed to participate in the project voluntarily. As the project did not involve any children or vulnerable adults, I did not need to comply with UK child protection clearance procedures.

3.7.2. Right to Withdraw

According to BERA (2011), participants have the right to know that they can withdraw from the research without providing a reason. It was ensured that all the participants involved in this project knew about their right to withdraw on the commencement of data collection. The contact details were also shared with them should they wish to express this right. They were informed that they could withdraw from the project up to four weeks after their engagement in the project. It was also emphasised that choosing not to be involved in the study would not affect them negatively in any way and any data already collected would be erased and would not be included in the final analyses.

3.7.3. Anonymity, Confidentiality and Privacy

BERA (2011) reminds researchers that participants' data should be treated in a way that it is confidential and anonymous. Prior to the commencement of data collection, the participants were informed that their data would be anonymised while writing up the project and it would not be shared with third parties.

This study also complied with the legal requirements in relation to the storage and use of personal data as set down by the Data Protection Act (1998, 2018). All the paper documents (e.g., consent forms, participant questionnaires) which included participant information were stored in a locked fillet cabinet which could be accessed by me only. Any information stored digitally (e.g. interview recordings) were password protected in my personal computer. Finally, in line with the Data Protection act (1998), any confidential information will be stored no longer than necessary and be safely destroyed in line with the requirements.

3.7.4. Openness and disclosure

It is clearly stated in BERA's (2011) guidelines that researchers can use deception only if it is a necessity of the research design and if they make sure that the welfare of the researchers is not put in jeopardy. If any deception is used, participants should be

informed about the nature at the earliest feasible opportunity. In this project, in Study 1 and Study 2, there was no deception. However, Study 3 was an experimental study and due to the nature of this study, some minor deception was used. There were three groups of participants and two of these groups were primed with a different negative stereotype regarding their gender and competency in language learning. The negative stereotypes were used to explore how such input could influence the participants' performance.

As there are several other studies adopting the same approach that have been ethically approved studies at Edge Hill University (e.g., Kaye & Pennington, 2016; Kaye, Pennington, & McCann, 2018; Pennington, Kaye, & McCann, 2018) and the instructions which were used in this study have previously been employed in peer-reviewed, published journal articles, it was decided that there were no apparent physical or psychological risks posed to participants (see Appendix 2, 3, 4 for all ethical materials for participants). However, as an experimenter, I was still observing the participants throughout the experiment and was ready to help them in case of any distress. As mentioned above, risk of harm to participants as a result of participating in this study was considered quite low, but it was assumed that the major problem, if any, would be that the input mentioned above could make the students feel anxious, worried or stressed. It was also possible that as a result of the input given, some students might perform poorly in the test. To reduce or eliminate these kinds of effects, all the participants were debriefed about the nature of the study once they had completed the test (Please see Appendix 3.4 for the Participant Debrief form). It was explained to the participants that the negative stereotypes did not apply to them as a person or were not true of their ability. It was made clear that such stereotypes were used to explore the effects that negative stereotypes might have on performance.

The type and level of threat was no more than students would encounter during their everyday lives. As mentioned in the literature review part, in most of the language learning environments, there is a commonly held belief that females are better than males in foreign language learning. In alignment with this argument, several studies have found that females tend to be more confident in language classes than males because it is stereotypically believed to be a female activity (Eccles et al., 1989; Pajares & Valiante, 1997; Pajares et al., 1999). This experiment just aimed to make this stereotype salient and assess the immediate effects of it on students' self-efficacy, anxiety and performance.

3.8. Chapter Summary

This chapter has described the research paradigm chosen and the philosophical assumptions it entails. It has also reviewed and rationalised the use of Mixed Methods Research and a convergent parallel design. A particular consideration has been given to the specific data collection methods, namely closed questionnaires, semi-structured interviews and experiments. Both advantages and disadvantages of using these methods have been covered and the extent to which the disadvantages were addressed in the current study has been explained. Furthermore, the data analysis processes for each study have been presented. The chapter has concluded with the summary of the ethical considerations that were relevant to this research.

CHAPTER 4: STUDY 1

4.0. Chapter Outline

As stated in Chapter 3, this programme of research consisted of three separate studies all of which had distinct but related aims and objectives. This chapter presents the findings of Study 1 which examined the relationship between foreign language (L2) learners' gender stereotyped beliefs pertaining to L2 learning, as well as to their L2 teachers' beliefs, and their L2 performance in light of two mediators, namely L2 anxiety and L2 self-efficacy. The chapter is composed of three main parts. First, Section 4.3 presents the pilot study which is concerned with the data collection instruments used in Study 1. To be able to answer the research questions presented in Section 4.2, three new questionnaires were developed for the purpose of the current research. These questionnaires have undergone robust development and validation procedures which are discussed in detail in Section 4.3. Secondly, Section 4.4 focuses on the main study. Using the data gained through the new questionnaires, the two structural equational models which were built and tested to determine the extent to which L2 learners' gender stereotyped beliefs pertaining to L2 learning, as well as to their L2 teachers' beliefs were linked to their L2 performance through L2 anxiety and L2 self-efficacy. Finally, the discussion gives a summary and critique of the findings. Limitations of the study and implications for further research are also provided.

4.1. Introduction

Inquiry into the role of individual differences (IDs) in language learning is not a new phenomenon in the field psychology of SLS (Dörnyei & Ryan, 2015). The direction of inquiry, however, has considerably changed over the last decade. As Dörnyei and Ryan (2015) point out, initial studies concerned with diverse characteristics within L2 learners held at least four erroneous assumptions that need to be challenged: 1) IDs are distinct from each other, 2) they are stable over time and across situations, 3) the components of each ID are only moderately related to each other, and 4) they are independent from the external factors (e.g., culture) surrounding them. Being informed by the fact that context and time are indeed crucial in achieving a more comprehensive understanding of IDs and the extent to which they influence L2 language learners and their performance, studies in the psychology of language learning have adopted a more dynamic perspective which suggests that IDs interact with each other, and they are socially interdependent (Dörnyei & Ryan, 2015). In line with this assertion, the current study examines a specific cultural and social context of higher education (HE) in Turkey, rather than assuming that IDs play similar roles in language learning universally. In this

chosen context, the aim is to investigate the extent to which gender stereotyping of L2 learning relates to Turkish L2 learners' language performance through the mediating roles of L2 anxiety and L2 self-efficacy.

L2 anxiety is among the psychological constructs which have been intensively studied over the past three decades (Dewaele, 2017). The debilitating effect of L2 anxiety on L2 learning has been supported by a wealth of empirical evidence gained through studies done with different kinds of language learners from a range of countries (Dewaele & MacIntyre, 2014; Dewaele, 2017; Horwitz et al., 1986; Horwitz & Young, 1991; MacIntyre, 2017; Phillips, 1992; Saito & Samimy, 1996). However, these studies have been limited in that they have mainly concentrated on the direct relationship between L2 anxiety and performance until recently (Dörnyei & Ryan, 2015; MacIntyre, 2017; Pavlenko, 2013). According to Pavlenko (2013), a more contextualised approach is needed because psychological factors such as motivation, anxiety and willingness to communicate do not stand alone. These are continuously interacting with many other personal (e.g., age and aptitude), social (e.g., ethnicity and culture), and situational factors (e.g., classroom environment and peers) as well as with each other (MacIntyre, 2017). In line with these suggestions, the 2010s have seen the rapid development of a more contextualised way of studying L2 anxiety which is called dynamic approach. Researchers have showed an increased interest in the interactions between L2 anxiety and other factors such as enjoyment and self-efficacy (Dewaele, 2013; Dewaele & MacIntyre, 2014; Gregersen et al., 2014; MacIntyre, 2012; MacIntyre & Serroul, 2015; Mills et al., 2006). The current research is, therefore, timely and needed as it adopts a dynamic approach to studying IDs such as self-efficacy and L2 anxiety. Rather than investigating the direct link between a specific ID and L2 performance, the current study provides a more contextualised picture of the IDs by examining L2 anxiety and L2 self-efficacy in relation to gender stereotyping of L2 learning in a particular context, namely Turkey.

Even though there is ample literature on both L2 anxiety and L2 self-efficacy and the extent to which they impact L2 performance, little empirical evidence is available on the link between gender stereotyping of L2 learning and these two constructs. In most foreign language learning environments, there is a commonly held belief that language learning is a feminine domain (Carr & Pauwels, 2006; Pomerantz, 2008; Schmenk, 2004). It is hypothesised that such gender stereotyping of academic domains is problematic because how an academic subject is perceived by females and males might influence their achievement related perceptions such as self-efficacy (Pajares & Schunk, 2001). In

alignment with this argument, several studies have found that females tend to have higher self-efficacy beliefs in language studies compared to males (see Wang et al., 2014). This raises the concern that there might be gender differences in L2 achievement since it is well-documented in the literature that self-related perceptions such as self-efficacy affect L2 performance in various ways. For example, learners with higher self-efficacy beliefs are reported to achieve higher L2 performance scores (Hsieh & Kang, 2010); have lower L2 anxiety (Mills et al., 2006) and use L2 learning strategies more effectively (Graham, 2007; Graham & Macaro, 2008). Also, while students with low L2 self-efficacy may tend to spend more time on simple and straightforward tasks with minimal effort and patience, students with higher self-efficacy beliefs tend to be more willing and put more effort when it comes to challenging tasks (Mills et al., 2006).

Given that self-efficacy is a key asset to L2 achievement, it is necessary to develop an understanding of the role of gender stereotyping of L2 learning in L2 learners' self-efficacy and achievement. However, to date, research concerning gender stereotyping of academic subjects has mainly focused on females' beliefs and their underachievement in the so-called masculine subjects such as maths (Plante et al., 2009). Especially in the last decade, far too little attention has been given to males' beliefs and their achievement in the academic fields such as language studies which are mostly associated with females. Thus, one of the aims of the current study is to address this limitation by providing a contemporary picture of the issue of gender stereotyping in L2 learning. This is achieved through exploring the link between both female and male L2 learners' gender stereotyped beliefs and their L2 achievement via the mediating role of L2 anxiety and L2 self-efficacy.

Although researchers seem to show that stereotypical beliefs pertaining to gender are mainly similar to each other, gender and how it is perceived by people change across cultures, time and different communities of practice (Schmenk, 2004). That is, perceptions about gender and gender stereotypes are not applicable to all situations. It is important to keep in mind that gender cannot be isolated from other aspects of social identity and relations, that it does not have the same meaning across communities, and that the linguistic usage of its regarded meaning is not the same across communities (Eckert & McConnell-Ginet, 1992). Therefore, while acknowledging the studies concerned with gender stereotypes and their effects, the current study contributes to the literature by targeting an academic-based community in Turkey where there are clear culturally situated distinctions between femininity and masculinity. To date, it has not been explored whether such stereotypes in respect of L2 learning exist among Turkish

EFL learners at university level in Turkey and to what extent these beliefs influence L2 learners' performance via the mediating roles of L2 anxiety and L2 self-efficacy. To address this gap in the literature, a number of aims and objectives were devised for Study 1 which are provided below.

4.2. Research Aims and Questions of Study 1

Study 1 quantitatively assessed the impact of participants' gender stereotype endorsement level and what they thought their teachers believed on their L2 performance through the mediating roles of L2 self-efficacy and L2 anxiety. It aimed to ascertain whether L2 learners in the chosen context, namely Turkish university, endorsed the common gender stereotypes in foreign language learning and the extent to which these beliefs indirectly related to subsequent L2 performance mediated by L2 anxiety and L2 self-efficacy.

The study addressed the following research questions and the sub-questions:

1. To what extent does learners' level of gender stereotype endorsement about L2 learning relate to their L2 performance through the mediating role of L2 anxiety and L2 self-efficacy?
 - 3.1. To what extent does L2 anxiety mediate the link between gender stereotyped beliefs and L2 performance?
 - 3.2. To what extent does L2 self-efficacy mediate the link between gender stereotyped beliefs and L2 performance?

4.3. Pilot Study

Pilot studies, also known as feasibility studies, are trial runs of comprehensive research studies. They can also be conducted to pre-test a particular data collection tool such as a questionnaire or interview questions (Van Teijlingen & Hundley, 2001). Pilot studies are regarded as a crucial component in the data collection process of research studies (Van Teijlingen & Hundley, 2001). As Blaxter et al. (2010) emphasise, researchers need pilot studies because even if they plan every detail of their research process well in advance, in the real study, their plans might need to change due to unexpected circumstances. Piloting is the key element that prevents these unexpected problems in the process of data collection and provides researchers with "re-assessment without tears" (Blaxter et al., 2010, p.138). It enables researchers to try out the research techniques and methods which are designed for the purpose of a particular study, determine whether and to what extent these work in practice, and adjust them when necessary (Blaxter et al., 2010). However, it is important to keep in mind that piloting a

study does not guarantee that the main study will be completely reliable, but it makes it stronger (Van Teijlingen & Hundley, 2001).

Considering that piloting might reduce the risk of possible failures in the main study, a pilot study was conducted in the current research. Several researchers agree that pilot studies have a variety of functions such as increasing the reliability, validity and practicability of the questionnaire (Oppenheim, 1992; Morrison, 1993; Wilson & McLean 1994; Cohen et al., 1997). Based on the functions summarised by Cohen et al. (2018), piloting was carried out in Study 1 to:

- validate the Multidimensional Language Class Anxiety Scale (MLCAS), the Questionnaire of Self-Efficacy Beliefs in Learning a New Language (QSL) and the Questionnaire of Language Learners' Gender Stereotyped Beliefs (QLGB)
- check the clarity of the questionnaire items, instructions and layout in the above-mentioned questionnaires,
- check readability levels for the participants and identify commonly misunderstood or non-completed items,
- check any items that are ambiguous or have a problem in wording,
- identify which items are too easy, too difficult, too complex or too remote from the participants' experience,
- check how long it takes to complete all the three questionnaires and see whether they are too long to complete in a limited time.
- practice the coding/classification system for data analysis.

4.3.1. Back-Translation

As the target samples consisted of Turkish students learning EFL, the Turkish version of the questionnaires were used in the current study. To increase the accuracy of the questionnaires, back-translation method was used. Back translation is the process of translating a text from the target language back to the source language (Brislin, 1970; McDermott & Palchane, 1994). It is suggested that at least two bilingual people should be involved in this process (McDermott & Palchane, 1994). One of these people translates the text from the source to the target language. Then, the second bilingual person takes the translated version of the text and blindly back-translates it from the target language to the source language. This enables researchers to have two versions of the original text for comparison (McDermott & Palchane, 1994).

For the purpose of the current study, the initial version of the scales which were in English was translated into Turkish by two different certified translators. The Turkish versions were then back-translated into English by two different translators who were not aware of the original version of the scale. All the versions were compared for consistency and accuracy by the experts. As the original and translated versions were all along the same lines, one final Turkish version for each questionnaire was created based on the translated versions. To establish the face validity, the final versions of the scales were piloted with five Turkish native speakers. They were checked for clarity and accuracy. Minor changes of wording in items were operated to avoid errors that might lead to differences in meaning.

4.3.2. Participants

The pilot study was done among Turkish adult learners of English in Turkey. As the current research concerned L2 learners studying English as a foreign language, it was a necessity to find participants within this context. As outlined in the Literature Review in Chapter 2, learning a foreign language is distinct from learning a second language in that people learning a second language are generally exposed to the target language in their daily lives. However, foreign language learners have much more limited opportunities in terms of using the target language as a communication tool (please see Section 2.1.3 for a detailed description of the concepts). In particular, it was aimed to see the effect of Turkish community and L2 teachers who are Turkish on Turkish learners' foreign language learning experience. Therefore, it was decided to do the pilot study in Turkey as well as the main study. A total of 323 participants (male = 176, 45.5%; female = 147, 54.5%) with a mean age of 18.85 years ($SD = 1.3$) participated in the pilot study. There were no missing data regarding the variables needed for the study.

4.3.3. Data collection Instruments and Procedures

The data was collected using three questionnaires all of which were developed for the purpose of the current study and underwent a number of validation processes. All the questionnaires used in the pilot study are provided in Appendix 4.1. The following sections are devoted to these development and validation procedures of the data collection instruments used. The section is organised in two parts. The first part outlines the structure of the questionnaires and describes how the initial items were developed for each of them. The second part presents the results gained from the analytic procedures, namely Exploratory Factor Analysis and Confirmatory Factor Analysis.

4.3.3.1. Background Questionnaire and the Questionnaires of Language Learners' Gender Stereotyped Beliefs

As one of the main aims of this PhD thesis was to explore whether language learners have any gender stereotyped beliefs pertaining to foreign language learning and to what extent these beliefs affect their performance through the mediating role of self-efficacy and anxiety, a gender stereotyped beliefs questionnaire was needed. As there has not been a comprehensive questionnaire addressing this need to date, a new questionnaire which was mainly informed by the framework developed by Whitehead (1996) and Schmader, Johns and Barquissau (2004) was devised. The questionnaire consists of three parts:

Part 1: Background information

The main purpose of this section was to gather information about participants' background which would help characterise the sample. The questions asked about gender (i.e., male or female), age, home region (i.e., Black Sea Region, Marmara Region, Aegean Region, Mediterranean Region, Central Anatolia Region, Eastern Anatolia Region, South-eastern Anatolia Region), duration of English study, any other languages spoken, and time spent abroad (see Appendix 4.1.1 for more details).

Part 2: Learners' Gender Stereotyped Perceptions of Academic Subjects

The second part of the questionnaire was adapted from Whitehead (1996) (Appendix 4.1.2). Participants were given 11 academic subjects such as maths, engineering and nursing and expected to rate them using a 5-item scale (1 - Females better, 5 - Males better) with a neutral value of 3 (Both sexes are equally better). This section aimed to explore whether participants had gender stereotyped beliefs about certain academic subjects among which is English. Although the main focus of this PhD project was foreign language learning only, it was decided that enquiring into participants' beliefs about the other subjects as well would be beneficial. By providing a list of academic subjects, participants' attention was not drawn to foreign language learning particularly which might create priming effect. That is, participants' responses were not affected by making the intention of the study salient at this stage.

Part 3: Questionnaire of Learners' Gender Stereotyped Perceptions of L2 learning

The last part of the questionnaire consisted of eight items. Four of these items (e.g., I think language learning is a feminine domain) aimed to assess language learners' gender stereotyped beliefs about foreign language learning. As the role of teachers as an agent of socialisation was part of the research as well, the other four items e.g., my

lecturers think that females have more ability in language learning than males.) were about students' perceptions about their language teachers' gender stereotyped beliefs. The participants were asked to rate these statements using the 5-point response scale (1-Strongly Disagree-5-Strongly Agree) (Appendix 4.1.3).

4.3.3.2. Multidimensional Language Class Anxiety Scale (MCLAS)

Within the field of SLS, L2 anxiety has been one of the primary concerns of researchers and practitioners for more than three decades (Dörnyei & Ryan, 2015). There is now accumulating evidence indicating that L2 anxiety is one of the most negatively influential psychological variables in L2 learning process (Horwitz, 2001; Horwitz, 2017; Park & Lee, 2005). Considering previous research has also consistently reported that L2 anxiety has been experienced by many L2 learners (Horwitz et al., 1986; Horwitz, 2000; Lu & Liu, 2011; Tsui, 1996), it is crucial for researchers to work towards developing a complete understanding of L2 anxiety adopting a variety of perspectives and approaches. According to Schutz and Pekrun (2007), as it not always easy and straightforward to understand the role of emotions in the learning processes, investigations into emotions such as anxiety should be theory-driven. In response to this call, this study developed a new theory-driven measure, the Multidimensional Language Class Anxiety Scale (MLCAS), which constitutes five brief scales concerning the major necessary components of L2 teaching and learning, namely listening, reading, writing, speaking and testing and presents evidence of reliability and validity for the data gathered using the MLCAS.

The MLCAS was developed using items adapted from the Achievement Emotions Questionnaire (AEQ) (Pekrun, Goetz, Frenzel, Barchfeld, & Perry, 2011). The AEQ consisted of four positive emotions (enjoyment, hope, pride, and relief) and five negative emotions (anger, anxiety, hopelessness, shame, and boredom). Only anxiety related items, however, were considered in the current research. Pekrun et al. (2011) contextualise emotions as class-related, learning-related, and test-related emotions. Items were chosen from among the class-related anxiety items because this study focuses on L2 anxiety experienced by EFL learners and EFL learning generally takes place in contexts where the target language is not spoken outside the classroom (Tavakoli, 2013). As most of learning takes place in classroom environment, the class-related items of the AEQ were more suitable for the overall purpose of this study than the learning related items. Cronbach alpha was reported as $\alpha = .86$ for the class-related anxiety sub-scale of the AEQ (Pekrun, Götz, & Perry, 2005; Pekrun et al., 2011).

Pekrun et al. (2011) propose that emotions consist of psychological processes which are interconnected to each other. These processes include affective (e.g., feeling tense and uneasy), cognitive (e.g., worrying), physiological (e.g., being activated physiologically), and motivational (e.g., avoidance) components (Pekrun et al., 2011). Therefore, for each emotion, they include affective, cognitive, physiological, and motivational components. In this study, only the affective, cognitive and physiological components were considered. The motivational component was excluded to avoid construct overlap. Motivation is widely considered to be a cognate, but distinct construct, to that of anxiety (Cheng et al., 2014; Gardner, Day, & MacIntyre, 1992; Hagtvet & Benson, 1997).

Table 4.1. The original class-related items created by Pekrun et al. (2005)

Components	Class-related Items
Affective	Thinking about class makes me feel uneasy I feel nervous in class.
Cognitive	Even before class, I worry whether I will be able to understand the material. I worry whether I'm sufficiently prepared for the lesson.
Physiological	I get tense in class. When I don't understand something important in class, my heart races.

The most suitable 2 items pertaining to cognitive, affective and physiological components were chosen on the basis of face validity. Taking this into consideration, preference was given to the items that were easy to adapt to L2 learning situations and that would be translated into Turkish, the language of the target sample, without losing their meanings. The original class-related items created by Pekrun et al. (2005) are presented in Table 4.1. The full version of the questionnaire used in the pilot study is provided in Appendix 4.2. The adapted version of the items which were used in the main study is also provided below in Table 4.9.

Each item was adapted for each language learning skill as seen in Table 4.1. The procedure was repeated for all the skill-based anxieties and test anxiety and for three components of anxiety (i.e., affective, cognitive, physiological). As a result of this procedure, a 3x5 dimensional scale emerged (see Table 4.2).

Table 4.2. The Structure of the MLCAS

	LAA	SAA	RAA	WAA	CTA	Total
Affective	2 items	2 items	2 items	2 items	2 items	10
Cognitive	2 items	2 items	2 items	2 items	2 items	10
Physiological	2 items	2 items	2 items	2 items	2 items	10
Total	6	6	6	6	6	30

Note: LLA = Listening Activity Anxiety, SAA = Speaking Activity Anxiety, RAA = Reading Anxiety Activity, WAA = Writing Activity Anxiety, CTA = Classroom Testing Anxiety

4.3.3.3. Questionnaire of Self-Efficacy Beliefs in Learning a New Language (QSLL)

As Bandura (2006) emphasises, studies concerning perceived self-efficacy beliefs need to avoid 'one-size-fits-all' approach when looking for appropriate measures. This kind of approach is limited in terms of explaining and predicting efficacy domains because a scale that is constructed for one purpose may have little or no relevance to another purpose (Bandura, 2006). Using such insufficient and inappropriate scales may create ambiguous results (Bandura, 2006). To avoid such difficulties, self-efficacy scales must be tailored to a particular context, demands and tasks. Therefore, although there are some self-efficacy questionnaires used in the previous studies (e.g., Mills et al., 2006; Wang et al., 2013) and the data gathered using these scales are reported to be valid and reliable, a new questionnaire which is in line with the purpose of the current research was needed for the current study.

The QSLL was created based on Bandura's (2006) guidelines which are designed to help researchers aiming to construct a self-efficacy measure. These guidelines address the thorny issues concerning content validity of self-efficacy scales, domain specification, and gradations of challenge all of which are discussed in Chapter 2 in Section 2.3.2.2. The questionnaire items were designed based on the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR) by Council of Europe (2001). The CEFR (2001) is an international standard for describing language ability. It aims to provide a common basis for anything related to language learning such as language syllabuses, curriculum guidelines, examinations and textbooks across Europe. It outlines the knowledge and skills that each language learners need to have in order to use the language effectively (Council of Europe, 2001). According to this

comprehensive framework, there are six broad levels that language learners can achieve. These are illustrated in Figure 3.5 in Chapter 3.

CEFR (2001) also provides a detailed global scale for each level (see Table 3.2 in Chapter 3 for further details). According to the CEFR (2001), "Such a simple 'global' representation will make it easier to communicate the system to non-specialist users and will also provide teachers and curriculum planners with orientation points." (p.24). In addition to this global scale, the CEFR (2001) presents a more detailed grid called Common Reference Levels: self-assessment grid (Appendix 4.3). In this grid, there are 'can do' descriptors applied to aforementioned six levels and these levels are analysed under three categories which are understanding (listening and reading), speaking (spoken interaction, spoken production) and writing. The items of the QSLL are created based on these 'can do' descriptors. For example, according to this grid, a language learner who completes C1 level is able to say, "I can understand television programmes and films without too much effort". Based on this descriptor, I created the item: "I can understand English TV news programs without English/Turkish subtitles". This procedure was repeated for all the items. As suggested by Bandura (2006), the items were chosen in a way that they represent gradations of challenges. A total of 20 items (5 items for each language skill) which were scored from 1 (Strongly Disagree) to 5 (Strongly Agree) were generated (Appendix 4.4). Broadly speaking, all the items aimed to find out whether the students believe that they can perform a specific task attributed to one of four language skills at one of six levels.

It is suggested that a self-efficacy scale with the 0-100 response format is a stronger predictor of performance than one with a 5-point scale (Pajares, Hartley, & Valiante, 2001). Bandura (2006) also suggests that scales should not use only a few steps (i.e. 5-point scale) because people usually avoid expressing their opinions using the extreme poles. This makes scales less sensitive and less reliable (Bandura, 2006). However, research has found little or no statistical difference between 5-, 7-, or 10-point scales in terms of their effect on the outcome of experiments (see Dawes, 2008). In his empirical research, Dawes (2008) confirms that "none of the three formats is less desirable from the viewpoint of obtaining data that will be used for regression analysis" (p.75). Also, Miller (1956), one of the most highly cited papers in psychology (Gorenflo & McConnell, 1991), explained that the human memory has a limited capacity and 7 (plus-minus two) would be ideal when designing scales. Based on the above suggestions and the fact that the other two questionnaires were also designed using 5-point response scale (1-

Strongly Disagree-5-Strongly Agree), the items in this questionnaire were rated on a 5-point Likert scale.

4.3.4. Results

4.3.4.1. Background Questionnaire and the Questionnaires of Language Learners' Gender Stereotyped Beliefs

Table 4.3. Frequency analysis of demographics and foreign language background of the participants

Demographic		N	Percentage
Gender	Male	176	45.5%
	Female	147	54.5%
Age	<18 years	5	1.5%
	18-25 Years	310	96%
	>25 years	1	0.3%
	Missing	7	2.2%
Home region	Black Sea Region	82	25.4%
	Marmara Region	77	23.8%
	Aegean Region	27	8.4%
	Mediterranean Region	45	13.9%
	Central Anatolia Region	45	13.9%
	South-eastern Anatolia Region	16	5.0%
	Eastern Anatolia Region	29	9.0%
	Missing	2	0.6%
Duration of English study	< 6 months	288	89.1%
	6-12 months	5	1.5%
	> 12 months	30	9.3%
Foreign Language other than English	No	216	66.9%
	Yes	107	33.1%
Been abroad	No	234	72.4%
	Yes	89	27.6%

Part 1. Descriptive Results for Background Information

Descriptive results concerning the participants' background information is provided in Table 4.3. As seen in Table 4.3., the participants who were recruited from three (1 state, 2 private) universities based in Istanbul, Turkey were originally coming from different

geographic and cultural backgrounds: Black Sea Region = 82 (25.5%); Marmara Region = 77 (24%); Aegean Region = 27 (8.4%); Mediterranean Region = 45 (14%); Central Anatolia Region = 45 (14%); South-eastern Region = 16 (5%); and Eastern Anatolia Region = 29 (9%).

The majority of the participants (72.4%) had not been abroad before meaning the students did not have frequent contact with native speakers of English. Also, learning English as a foreign language was the first L2 experience for 66.9% of the participants as they had not studied any other foreign languages than English before.

Part 2. Learners' Gender Stereotyped Perceptions of Academic Subjects

Table 4.4. provides the descriptive results gained from the participants' gender stereotyped perceptions of a number of academic subjects. To determine whether the participants had gender stereotypical beliefs about the subjects provided, a series of one sample t-tests were conducted for females and males separately to establish the extent to which each sub-sample independently endorsed stereotypical attitudes from the neutral point. On the scale used, a midpoint of 3 (both females and males are equally better) was present for neutral. As such, any means under 3 would suggest female association and any over 3 as male association.

The results indicated that females endorsed the gender stereotype that more females than males were better at Social Sciences: $t(146) = -8.22, p < .001$, Performing Arts: $t(146) = -5.69, p < .001$, Arts Subjects: $t(143) = -2.60, p < .001$, English: $t(144) = -2.90, p < .01$, Biological Sciences: $t(143) = -2.61, p < .01$ and Practical Subjects: $t(146) = -15.54, p < .001$. However, they endorsed that more males than females were better at Physical Sciences: $t(144) = 5.03, p < .001$, Economics and related subjects: $t(146) = 7.32, p < .001$, Mathematics and related subjects: $t(145) = 5.47, p < .001$ and Engineering: $t(146) = 9.20, p < .001$. They did not significantly endorse gender-subject competence stereotypes for Geography and related subjects: $t(146) = .92, p > .05$.

With regards to males, they also endorsed that females were better than males at Social Sciences: $t(173) = -6.54, p < .001$, Performing Arts: $t(173) = -5.82, p < .001$, Arts Subjects: $t(173) = -3.24, p < .001$, English: $t(174) = -3.30, p < .01$, and Practical Subjects: $t(174) = -16.96, p < .001$. Similar to females, males also believed that more males than females were better at Physical Sciences: $t(173) = 10.43, p < .001$, Economics and related subjects: $t(173) = 12.81, p < .001$, Mathematics and related subjects: $t(173) =$

11.14, $p < .001$ and Engineering: $t(173) = 23.57$, $p < .001$. The male participants did not significantly endorse gender-subject competence stereotypes for Geography and related subjects: $t(173) = 1.51$, $p > .05$ and Biological sciences: $t(174) = 1.46$, $p > .05$. As seen in the results, both females and males had gender stereotypical beliefs concerning English.

Table 4.4. Descriptive Statistics of Learners' gender Stereotyped Perceptions of Subjects - The Pilot Study

	N	Possible Range	Observed Range	M	SD	Skewness	Kurtosis
1. Social Sciences	321	1-5	1-5	2.53	.822	.177	.629
2. Physical Sciences	319	1-5	1-5	3.46	.742	.460	.555
3. Performing Arts	321	1-5	1-5	2.67	.730	-.374	1.620
4. Geography and related subjects	321	1-5	1-5	3.07	.695	.357	2.179
5. Arts subjects	318	1-5	1-5	2.81	.794	.154	1.487
6. English	320	1-5	1-5	2.85	.598	-.202	4.513
7. Biological Sciences	319	1-5	1-5	2.98	.735	.221	1.772
8. Economics and related subjects	321	1-5	1-5	3.68	.852	-.179	.152
9. Mathematics and related subjects	320	1-5	1-5	3.50	.755	.417	.377
10. Practical subjects	322	1-5	1-5	1.97	.810	.629	.747
11. Engineering	323	1-5	1-5	3.98	.834	-.515	.352

It is important to note here that while filling in the questionnaires, some participants commented on the academic subjects given (not including English) and declared that they were too broad to rate. Therefore, it was decided that more specific examples were provided in the main study. The updated questionnaire is given in Appendix 4.5.

Part 3. Questionnaire of Learners' Gender Stereotyped Perceptions of L2 learning

As this Questionnaire of Learners' Gender Stereotyped Perceptions of L2 learning (QLGB) along with the MLCAS and QSLI were new and reliability and validity of data gathered using the new scales was not established before, a couple of analyses were conducted to validate the data gained using these questionnaires. It is recommended that scale development and validation studies should combine Exploratory Factor

Analysis (EFA) and Confirmatory Factor analysis (CFA) to strengthen the robustness of the development and validation procedure (Fabrigar, Wegener, MacCallum, & Strahan, 1999). In line with this suggestion, the pilot study data was used to pre-test the items of the questionnaires and determine its factor structure running an Exploratory Factor Analysis (EFA) in SPSS v.24. This was followed by a CFA for which the main study data. CFA aimed to confirm the last version of the scales and the results gained via the EFA using a CFA in *Mplus* 7.4 (Muthen & Muthen, 2013).

EFA. To assess the factor structure and psychometric properties of the QLGB, an EFA was performed in SPSS v.24 using the pilot study data. A series of statistics such as the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and Bartlett's test of sphericity were used in determining whether data analysis procedures were advisable. The factors were extracted from the pilot study data via Principal Axis Factoring with the direct oblimin rotation with Kaiser Normalization (an oblique rotation method assuming factors are correlated). The EFA results were evaluated to make decisions regarding the number of factors and the items corresponding to these factors. the. Following the EFA, I tested internal consistency reliability for the identified factors and their associated items.

Results. The KMO measure was .840 showing that the sample size was adequate for EFA. The Bartlett's test of sphericity was statistically significant ($p < .001$) which demonstrated that all the variables were correlated and can be part of the same factor. Convergence was achieved after five iterations, resulting in two factors with eigenvalues > 1 . The first factor which included five of the items accounted for 46.76%. The second factor consisted of two items which accounted for only 6.74%. The item "I don't think my lecturers find females more successful than males in the classroom." did not load on any of the factors (Table 4.5).

The EFA results were problematic for two reasons. First, it did not support the hypothesised factor structure of the scale. The scale originally aimed to measure two constructs: participants' gender stereotyped beliefs about L2 learning (Item 1, 3, 4, 5) and their perceptions about their teachers' gender stereotyped beliefs (Item 2, 6, 7, 8). However, an item (Item 2) which was supposed to be related to participants' perceptions about their L2 teachers' gender stereotyped beliefs loaded on the factor concerned with the participants' own gender stereotyped beliefs. This violates the assumption that all the items that represent participants' gender stereotyped beliefs about L2 learning constituted one factor and all the other items that assess the participants' perceptions

about their L2 teachers' gender stereotyped beliefs formed another factor. As such, the factor structure the EFA suggested had no theoretical base.

Table 4.5. Pattern Matrix for the QSLL

Items	Pattern coefficients	
	Factor 1	Factor 2
1. I think L2 learning is a feminine domain.	.785	
2. According to my lecturers, females perform better than males in L2 classrooms.	.470	
3. I think females are more talented in L2 learning than males.	.976	
4. Females are usually more successful in L2 learning than males	.753	
5. I don't think there are any gender differences in L2 learning ability.	.483	
6. My lecturers think that L2 learning is more suitable for females.		.676
7. My lecturers think that females are more talented in L2 learning than males.		.979
8. I don't think my lecturers find females more successful than males in the classroom	-	-

Note: Extraction Method: Principal Axis Factoring. Rotation Method: Direct Oblimin with Kaiser Normalization. Rotation converged in 5 iterations

Secondly, due to Item 2 loading on the first factor and Item 8 not loading on its purported factor, the second factor which is concerned with the participants' perceptions about their teachers' gender stereotyped beliefs had only two items. It is suggested that having more than two items is better in exploratory research in that this could enhance construct validity by increasing the chance of identifying the construct of interest sufficiently (Eisinga, Te Grotenhuis, & Pelzer, 2013). Therefore, it was decided that the factor structure suggested by EFA was not acceptable and the scale needed to be revised for the main study. The revised version of the scale can be seen below in Table 4.17.

Descriptive Statistics. The factor structure suggested by EFA was further scrutinised to analyse a number of statistics such as number of items in each construct, observed ranges, means, standard deviations, skewness, kurtosis of the subscales as well as the Cronbach's alpha (α). The descriptive statistics were generated using SPSS. As seen

from Table 4.6, Cronbach's alphas (α) were $\geq .81$ which exceeded .70 cut-off criterion for reliability (Nunnally & Bernstein, 1994) and suggested that the factors had internal consistency. The skewness and kurtosis statistics showed that the data was not normally distributed. It is clear from both the EFA and the descriptive analysis that the structure of the QLGB was different from the hypothesised one which sustained the need for a revised version of the QLGB.

Table 4.6. QLGB Statistics – The Pilot Study

	No. of Items	Possible Range	Observed Range	M	SD	Skewness	Kurtosis	Cronbach's α
Teacher Beliefs	2	2-10	1-10	2.97	1.51	1.814	3.657	.81
Learner Beliefs	5	5-25	4-25	8.25	4.02	1.616	2.576	.85

I also analysed the descriptive statistics of the individual items in each factor of the QLGB and compared them with the results gained from the item (i.e., English) given in the questionnaire of participants' gender stereotyped perceptions of academic subjects. The results are provided in Table 4.7. below.

The results showed that the means of the QLGB items ($M_s = 1.44 - 1.83$) were clearly much lower than the mean of English ($M = 2.85$). That is, the belief that females were better at L2 learning than males was endorsed more strongly in the QLGB than in the Questionnaire of Learners' Gender Stereotyped Perceptions of Academic Subjects. Such a result implied that the QLGB questionnaire items failed to abide by one of the conventional wisdom about optimal questionnaire design, namely avoiding leading questions that push respondents toward an answer (Krosnick, 2018). It is highly likely that the way the items were worded in the QLGB forced the respondents to provide an answer that did not accurately reflect their real opinion. Therefore, it was decided that items such as "I think L2 learning is a feminine domain" were avoided in the questionnaire in the main study and worded in a more neutral way. The revised version of the QLGB is discussed in detail below. The items are provided in Appendix 4.6.

Table 4.7. Descriptive Statistics of the individual items in QLGB

	N	Possible Range	Observed Range	M	SD	Skewness	Kurtosis
1. I think L2 learning is a feminine domain.	322	1-5	1-5	1.53	1.008	2.127	3.964
2. According to my lecturers, females perform better than males in L2 classrooms.	322	1-5	1-5	1.73	.983	1.264	.974
3. I think females are more talented in L2 learning than males.	323	1-5	1-5	1.83	1.104	1.296	.906
4. My lecturers think that L2 learning is more suitable for females.	322	1-5	1-5	1.48	.821	1.916	3.577
5. females are usually more successful in L2 learning than males	322	1-5	1-5	1.75	1.068	1.420	1.231
6. My lecturers think that females are more talented in L2 learning than males.	323	1-5	1-5	1.50	.824	1.753	2.917
7. I don't think there are any gender differences in L2 learning ability.	323	1-5	1-5	1.44	.915	2.332	4.932
8. English	320	1-5	1-5	2.85	.598	-.202	4.513

4.3.4.2. The Multidimensional Language Class Anxiety Scale

EFA. Separate EFAs for each focal construct (i.e., listening activity anxiety (LAA), reading activity anxiety (RAA), writing activity anxiety (WAA), speaking activity anxiety (SAA) and classroom testing anxiety (CTA) were run using the pilot study data. The reason for using separate EFAs for each subscale was method factors arising from similar item stem wording that EFA cannot account for. That is, the identical item content across domains would create large method factors. Therefore, the constructs were tested separately first which is why no cross-loadings were present in the results of this analysis. All the analyses were run in SPSS v.24. To determine whether data analysis procedures were advisable, a series of statistics such as the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and Bartlett's test of sphericity were utilised. The factors were extracted from the pilot study data via Principal Axis Factoring with the direct

oblimin rotation with Kaiser Normalization (an oblique rotation method assuming factors are correlated).

Results. The KMO measure was $> .87$ for each latent variable which indicated that the data were appropriate for factor analysis.

Table 4.8. Exploratory Factor Analysis Results – The Pilot Study Data

Items	Factors				
	RAA	WAA	LAA	SAA	CTA
1. Thinking about reading activities in class makes me feel uneasy.	0.892				
2. I feel nervous during reading activities in class.	0.885				
3. Even before reading activities in class, I worry whether I will be able to understand the material.	0.665				
4. I worry whether I am sufficiently prepared for reading activities in class	0.783				
5. I get tense during reading activities in class.	0.844				
6. When I don't understand something important during reading activities in class, my heart races.	0.722				
7. Thinking about writing activities in class makes me feel uneasy.		0.755			
8. I feel nervous during writing activities in class.		0.897			
9. Even before writing activities in class, I worry whether I will be able to understand the material.		0.714			
10. I worry whether I am sufficiently prepared for writing activities in class		0.696			
11. I get tense during writing activities in class.		0.588			
12. When I don't understand something important during writing activities in class, my heart races.		0.556			
13. Thinking about listening activities in class makes me feel uneasy.			0.879		

14. I feel nervous during listening activities in class.	0.877
15. Even before listening activities in class, I worry whether I will be able to understand the material.	0.774
16. I worry whether I am sufficiently prepared for listening activities in class	0.618
17. I get tense during listening activities in class.	0.910
18. When I don't understand something important during listening activities in class, my heart races.	0.543
19. Thinking about speaking activities in class makes me feel uneasy.	0.790
20. I feel nervous during speaking activities in class.	0.816
21. Even before speaking activities in class, I worry whether I will be able to understand the material.	0.770
22. I worry whether I am sufficiently prepared for speaking activities in class	0.791
23. I get tense during speaking activities in class.	0.845
24. When I don't understand something important during speaking activities in class, my heart races.	0.672
25. Thinking about taking an English language test makes me feel uneasy.	0.746
26. I feel nervous while taking an English language test.	0.919
27. Even before taking an English language test, I worry whether I will be able to understand the material.	0.509
28. I worry whether I am sufficiently prepared for taking an English language test.	0.764
29. I get tense while taking an English language test.	0.835
30. When I don't understand something important while taking English language test, my heart races.	0.697

Note: Listening Activity Anxiety (LAA), Reading Activity Anxiety (RAA), Writing Activity Anxiety (WAA), Speaking Activity Anxiety (SAA) and Classroom Testing Anxiety (CTA)

Bartlett's test of sphericity was also statistically significant ($p < .001$) suggesting that the correlation matrix was not an identity matrix. In other words, the variables tested were related and suitable for structure detection. Iterations needed to extract the resulting factors for LAA, SAA, WAA, RAA, CTA were 5, 5, 6, 5 and 6 respectively. The one-factor solution for each skill-based anxiety accounted for > 56% of the variance in the data for all variables. The pattern matrix (see Table 4.8) demonstrated that all items loaded onto their target factors ($\lambda \geq .51$).

Descriptive Statistics. SPSS v.24 was used to generate descriptive statistics. Number of items in each construct, observed ranges, means, standard deviations, skewness, kurtosis and internal consistency (Cronbach alpha) of the subscales are provided in Table 4.9. As shown, all parts of the MLCAS yielded a Cronbach alpha scores $\geq .87$ which meets the .70 cut-off criterion for reliability (Nunnally & Bernstein, 1994). The skewness and kurtosis statistics showed that all variables were normally distributed.

Table 4.9. Descriptive results of the domains of the MLCAS

	No. of Items	Possible Range	Observed Range	M	SD	Skewness	Kurtosis	Cronbach Alpha
LAA	6	6-30	6-30	14.43	5.61	.297	-.673	.89
SAA	6	6-30	6-30	15.25	5.95	.207	-.657	.90
RAA	6	6-30	6-30	12.93	5.50	.622	-.073	.91
WAA	6	6-30	6-29	14.20	5.17	.352	-.487	.87
CTA	6	6-30	6-30	17.12	5.96	-.057	-.718	.89

Note: Listening Activity Anxiety (LAA), Reading Activity Anxiety (RAA), Writing Activity Anxiety (WAA), Speaking Activity Anxiety (SAA), Classroom Testing Anxiety (CTA)

Based on participants' feedback on the scale, the wording of 'I get tense during listening activities' which is one of the physiological items and its equivalences (items for speaking, reading, writing and testing) were changed. The translated versions of the

words "worry" "get tense" and "feel nervous" are used interchangeably in Turkish. Although their meanings are not the same, not all people are aware of their nuances. Therefore, to make the differences clear, the item 'I get tense during listening activities' was changed into 'I get tense during listening activities (e.g., having tense muscles)'. By giving an example, it was made clear that this item is related to physiology. Except for this minor change, all the other items were kept the same for the main study.

4.3.4.3. The Questionnaire of Self-Efficacy Beliefs in Learning a New Language (QSLL)

EFA. The aforementioned EFA procedures were repeated for the Questionnaire of Self-Efficacy Beliefs in Learning a New Language (QSLL) as well. To determine the extent to which the statistics were advisable, KMO measure of sampling adequacy and Bartlett's test of sphericity were utilised. Principal Components extraction with the Promax (oblique) rotation with Kaiser Normalization (an oblique rotation method assuming factors are correlated) was adopted to extract the factors using the pilot study data. Based on the EFA results, the following steps were undertaken. First, factor loadings that were equal to or greater than .40 were retained. Second, any cross-loaded items were deleted. Third, any identified factors and items needed to be theoretically interpretable. Following the EFA, internal consistency reliability for the identified factors and their associated items was tested.

Results. The KMO measure was .922 which indicated that the data were appropriate for factor analysis. Bartlett's test of sphericity was statistically significant ($p < .001$) suggesting that the correlation matrix was not an identity matrix. In other words, the variables tested were related and suitable for structure detection. The initial EFA which required eight iterations to extract the resulting factors offered a three-factor solution with eigenvalues > 1 . The factors accounted for 54.3% of the total variance.

Further analysis on the factors revealed that Factor 1 corresponded to productive skills which are speaking and writing and Factor 3 to receptive skills namely listening and speaking. Factor 2, however, which consisted of 8 items (the items 1, 2, 6, 7, 8, 14, 16, 17), did not correspond to any particular skills or provide any other structure that was theoretically comprehensible. Also, one of the items (item 12) cross loaded on all the three factors. Such a result supported the removal of the items in Factor 2 and the cross-loading item and required us to run a second EFA.

Table 4.10. Pattern Matrix for the QSL

Items	Pattern coefficients	
	Factor 1	Factor 2
<i>Item 11.</i> I can discuss topics such as families, hobbies, work and travel with my classmates in English	.856	
<i>Item 13.</i> I can ask questions to my teacher and answer his/her questions in English.	.842	
<i>Item 18.</i> I can write a personal letter describing my experiences and impressions in English	.716	
<i>Item 19.</i> I can write an English essay giving reasons in support of or against a particular point of view.	.678	
<i>Item 20.</i> I can express myself in clear well-structured English text, expressing points of view at some length.	.651	
<i>Item 15.</i> I can express myself fluently and spontaneously without much obvious searching for expressions in English.	.614	
<i>Item 5.</i> I can understand English films without English/Turkish subtitles.		.916
<i>Item 4.</i> I can understand English TV news programs without English/Turkish subtitles.		.888
<i>Item 10.</i> I can read and understand long and complex factual and literary English texts (e.g., novels, articles, essays etc.).		.706
<i>Item 9.</i> I can read and understand English articles and reports concerned with contemporary problems		.528
<i>Item 3.</i> I can understand the main point of an English radio/TV program on a personal /professional interest.		.496

Note: Extraction Method: Principal Component Analysis. Rotation Method: Promax with Kaiser Normalization. Rotation converged in 3 iterations.

The second EFA was performed on the remaining 11 items using principal components analysis with Promax rotation once again. The matrix tests and other statistics supported the second EFA (KMO = .912, Bartlett's $p < .001$). Extraction of two factors was

supported by the eigenvalue > 1 criteria. Six items (three speaking items 11, 13, 15 and three writing items 18, 19, 24) loaded onto Factor 1 which I named production, and five items two reading items 9, 10 and three listening items 3, 4, 5) loaded onto Factor 2 which corresponded to reception. The two-factor solution accounted for 56% of the variance in the data. The pattern matrix (see Table 4.10) demonstrated that all items loaded onto their target factors and no items cross loaded ($\lambda > .4$).

Descriptive Statistics. Number of items in each construct, observed ranges, means, standard deviations, skewness, kurtosis of the subscales as well as the overall score for the general factors are provided in Table 4.11. Cronbach's alpha (α) was used to verify the internal consistency of the factors. As shown, both factors of the QSLL yielded Cronbach's alpha scores $\geq .80$ which meets the .70 cut-off criterion for reliability (Nunnally & Bernstein, 1994). The skewness and kurtosis statistics indicated that all variables were normally distributed.

Table 4.11. The QSLL Statistics – The Pilot Study

	No. of Items	Possible Range	Observed Range	M	SD	Skewness	Kurtosis	Cronbach's α
Reception	5	5-25	3-25	13.53	3.57	.136	.202	.80
Production	6	5-30	6-29	19.67	4.16	.149	-.273	.83

Based on students' feedback and queries during the pilot study, some items were reworded to make them more specific and contextualised which would make it easier for the participants to assess their language proficiency. The revised version of the QSLL is given in Appendix 4.7.

4.4. Main Study

4.4.1. Participants

It is suggested that participants who are involved in the piloting process should not be selected for the main study (Van Teijlingen & Hundley, 2002). Therefore, in the main study, 3 different (1 private, 2 state) universities which were again based in Istanbul,

Turkey were selected. Based on the convenience sampling method, 701 students (male = 49.6%; female = 50.6%) with a mean age of 19.17 years ($SD = 1.9$) were recruited. As in the pilot study, permission was gained from the directors of the Foreign Languages Schools and the teachers responsible for the selected classes. The students were also asked to give their individual consent on the first page of the questionnaire.

4.4.2. Data Collection Instruments

In the main study, the revised versions of the QLGB, the MLCAS and the QSLI were adopted. The data collected using these questionnaires were used in two ways. First, a CFA was run on the revised version of the questionnaires to confirm the results gained from the pilot study. Following the CFA, the main study data was utilised for the Structural Equation Modelling (SEM) analyses which are presented in a separate section. Participants' L2 performance was assessed using their average English proficiency test scores which they received at the end of the English preparatory programme. The maximum score that the participants could get was 100%. The scores were calculated by the universities themselves based on a number of quizzes, mid-term and end-year exams. Although each university which was involved in this study had their own assessment system, the structure of the exams they used to assess the English proficiency was similar across the universities. All the participants were assessed for their reading, writing, listening and speaking competencies which constituted one final score. The exams which the participants took during the English preparatory programme were prepared by a separate unit in each university, namely a testing office. The testing offices were composed of a number of experienced EFL teachers who were responsible for the content, preparation and implementation of the examinations to be held throughout the academic year. Each exam was double-marked internally, and the grades were moderated by a third marker when the first two markers could not agree. Overall, the maximum period of time between the participants' completion of questionnaires (in March - April, 2017) and the collection of the performance scores (June - July, 2017) was about five months.

4.4.3. Results

4.4.3.1. Background Questionnaire and the Questionnaire of Language Learners' Gender Stereotyped Beliefs

Part 1. Descriptive Results for Background Information

The descriptive results gained from the frequency analysis is presented in Table 4.12. The students were originally from Black Sea Region = 403, 57.5%; Marmara Region = 51, 7.3%; Aegean Region = 41, 5.8%; Mediterranean Region = 44, 6.3%; Central Anatolia

Region = 24, 3.4%; South-eastern Region = 10, 1.4%; and Eastern Anatolia Region = 27, 3.9%.

Table 4.12. Frequency analysis of demographics and foreign language background of the participants - The Main Study

Demographic		N	Percentage
Gender	Male	346	49.4%
	Female	355	50.6%
Age	<18 years	235	33.5%
	18-25 Years	435	62%
	>25 years	6	1%
	Missing	25	3.6. %
Home Region	Black Sea Region	51	7.3%
	Marmara Region	403	57.5%
	Aegean Region	41	5.8%
	Mediterranean Region	27	3.9%
	Central Anatolia Region	44	6.3%
	South-eastern Anatolia Region	10	1.4%
	Eastern Anatolia Region	24	3.4 %
	Missing	101	14.4%
Duration of English study at university	< 6 months	21	3%
	6-12 months	567	80.8%
	> 12 months	16	2.2%
	Missing	97	13.8%
Foreign Language other than English	No	457	65.1%
	Yes	154	33.1%
Been abroad	No	555	79.2%
	Yes	53	7.6%
	Missing	93	13.3%
English Study before university	< 12 months	8	1.1%
	1-5 years	244	34.8%
	5-10 years	163	23.3%
	> 10 years	177	25.2%
	Missing	109	15.5%
Quality of English Study before university	Definitely Enough	9	1.3%
	Enough	52	7.4%
	Not Enough	246	35.1%
	Definitely not Enough	297	42.4%
	Missing	97	13.8%

At the point of data collection, the 80.8% participants had been studying EFL for 6-12 months at university. From among the participants, only 7.6% of the participants had been abroad suggesting most of the participants had limited contact with English native speakers. Although most of the participants (83.3%) had studied English before the

university, 77.5% of them did not think that the EFL education was enough or of good quality.

Part 2. Learners' Gender Stereotyped Perceptions of Subjects

The descriptive results gained from the learners' gender stereotyped perceptions of subjects are provided in Table 4.13 below. As mentioned in the pilot study, the academic subjects presented were adapted based on the participants' feedback. As they indicated that some subjects were too broad to rate, some more specific subjects were added to the list. The participants in the main study were asked to rate a total of 16 academic subjects (it comprised 11 items in the pilot study) (see Table 4.13).

Table 4.13. Descriptive Statistics of Learners' gender Stereotyped Perceptions of Subjects

	N	Possible Range	Observed Range	M	SD	Skewness	Kurtosis
1. Fine arts.	610	1-5	1-5	2.90	.613	-.369	2.722
2. Educational Sciences.	610	1-5	1-5	2.86	.812	.300	.416
3. Sports Sciences.	612	1-5	1-5	3.88	.737	-.649	1.285
4. Aviation and Space Sciences.	612	1-5	1-5	3.85	.782	-.702	1.496
5. Architecture and Design.	612	1-5	1-5	2.60	.762	.145	.597
6. Economics and Administrative Sciences.	607	1-5	1-5	3.45	.725	.295	.091
7. Medical Sciences.	611	1-5	1-5	3.12	.679	.162	1.745
8. Dentistry.	608	1-5	1-5	3.03	.729	.213	.968
9. Health Sciences.	612	1-5	1-5	1.97	.758	.777	1.627
10. Life Sciences.	606	1-5	1-5	3.36	.658	.545	1.095
11. Humanities.	609	1-5	1-5	2.80	.736	.028	1.017
12. English.	610	1-5	1-5	2.72	.606	-.491	1.943
13. Engineering.	611	1-5	1-5	3.87	.703	-.294	.537
14. Communication.	612	1-5	1-5	2.66	.716	-.017	1.061
15. Tourism.	612	1-5	1-5	2.97	.777	-.162	.811
16. Agriculture.	612	1-5	1-5	3.50	.802	.047	.014

To establish the extent to which the participants had gender stereotypical beliefs about the subjects presented in Table 4.13, a number of one sample *t*-tests were run using SPSS v24. The analyses were run for females and males separately to establish the extent to which each sub-sample independently endorsed stereotypical attitudes from the neutral point. As discussed previously in the pilot study, a value of 3 was identified as the neutral point on the scale used (1= Females are always better, 5=Males are always better). As such, any means under 3 would suggest female association and any over 3 as male association.

The results demonstrated that females significantly endorsed the belief that females were better than males at Fine Arts: $t(302) = -4.10, p < .001$, Educational Sciences: $t(301) = -4.19, p < .001$, Architecture and Design: $t(303) = -12.07, p < .001$, Health Sciences: $t(303) = -25.02, p < .001$, Humanities: $t(301) = -6.13, p < .001$, English: $t(302) = -7.50, p < .001$, Communication: $t(303) = -7.23, p < .001$. In contrast, they endorsed that more males than females were better at Sport Sciences: $t(303) = 15.57, p < .001$, Aviation and Space Sciences: $t(303) = 19.01, p < .001$, Economics and Administrative Sciences: $t(301) = 9.01, p < .001$, Life Sciences: $t(301) = 6.95, p < .001$, Engineering: $t(302) = 17.95, p < .001$, Agriculture: $t(303) = 8.71, p < .001$. Females did not significantly endorse gender-subject competence stereotypes for Medical Sciences: $t(302) = 1.55, p > .05$, Dentistry: $t(301) = 1.31, p > .05$, Tourism: $t(303) = 1.52, p > .05$,

With regards to males, they endorsed the belief that females were better than males at Fine Arts: $t(306) = -1.98, p < .05$, Educational Sciences: $t(307) = -1.93, p = .05$, Architecture and Design: $t(307) = -6.68, p < .001$, Health Sciences: $t(307) = -22.70, p < .001$, Humanities: $t(306) = -3.24, p < .001$, English: $t(306) = -8.35, p < .001$, Communication: $t(307) = -9.26, p < .001$. Differently from females, males believed that females were better than males at Tourism: $t(307) = -2.48, p = .01$. In contrast, they endorsed that more males than females were better at Sport Sciences: $t(307) = 29.03, p < .001$, Aviation and Space Sciences: $t(307) = 19.19, p < .001$, Economics and Administrative Sciences: $t(304) = 12.66, p < .001$, Life Sciences: $t(303) = 12.06, p < .001$, Engineering: $t(307) = 26.67, p < .001$, Agriculture: $t(307) = 13.29, p < .001$. Unlike females who did not think there were any gender differences in Medical Sciences, males believed that males were better than females at Medical Sciences: $t(307) = 4.63, p > .05$. The only subject for which males did not significantly endorse gender-subject competence stereotypes was Dentistry: $t(305) = .76, p > .05$. Consistent with the pilot study data, the main study data also showed that both females and males held gender

stereotypical beliefs for English. That is, the participants significantly endorsed the belief that females were better at English than males.

Part 3. Questionnaire of Language Learners' Gender Stereotyped Beliefs

The revised version of the QLGB consisted of two sections. The first section which had five items concentrated exclusively on students' perceptions about learning English as a foreign language. The participants were provided with five items such as "Please indicate which gender is generally **good at** learning English." and asked to rate these statements using the scale below:

1	2	3	4	5
Always females	Often females	Sometimes Females Sometimes Males	Often Males	Always Males

In the second section, the same items were used to investigate participants' perceptions about their language teachers' gender stereotyped beliefs.

CFA. Being informed by the EFA results gained from the pilot study data, the adapted 10 items were modelled within a CFA using the main study data. In addition to the two-factor model (Figure 4.1), two alternative models (i.e., one factor model and a higher order factor model) were created to make a comparison between the models. The same goodness of fit indices were used to evaluate the model fit.

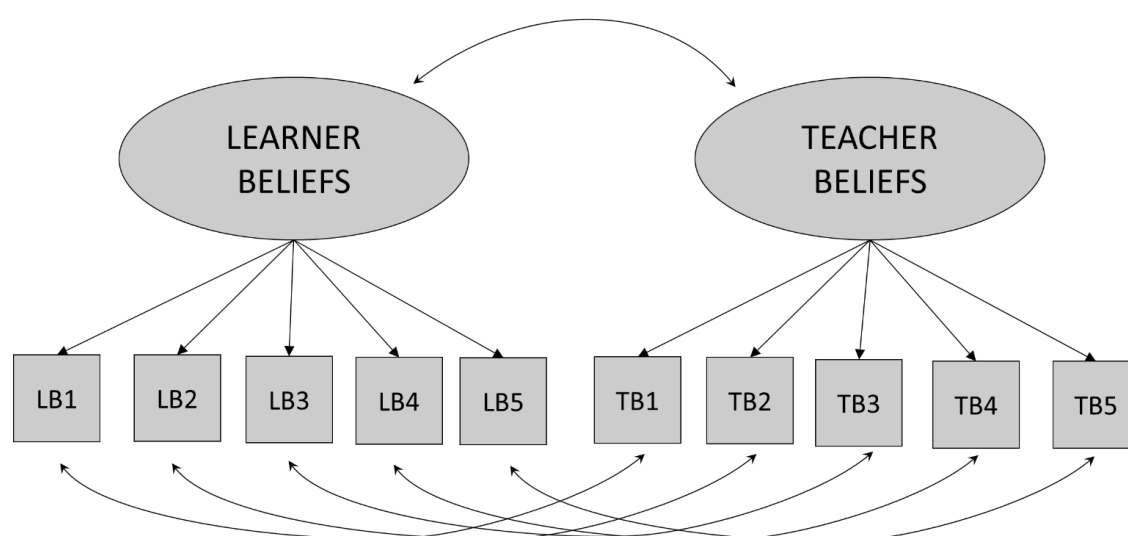


Figure 4.1. *Two-factor model of QLGB*

A number of model-fit criteria were used to evaluate the acceptability of all analysed models. The fit indices used to assess the models were the chi square (χ^2) statistic, degrees of freedom (*df*), the Root Mean Square Error of Approximation (RMSEA), the Standardised Root Mean Square Residual (SRMR), Comparative Fit Index (CFI), Tucker-Lewis index (TLI), Akaike Information Criterion (AIC) and sample-size adjusted Bayesian Information Criterion (aBIC). According to Hu and Bentler (1999), a good model is indicated by RMSEA < .05, SRMR < .08 and CFI and TLI > .95. As for AIC and aBIC, it is recommended that the model with the smallest value should be preferred (Hu & Bentler, 1999).

Table 4.14. Standardised Loadings for the Two-factor Solution – The Main Study

Items	Learner Perceptions	Teacher Perceptions
1. Please indicate which gender is generally good at learning English.	0.677	
2. Please indicate which gender is eager and motivated to learn English.	0.706	
3. Please indicate which gender is good at using the methods and strategies that are effective for learning English.	0.496	
4. Please indicate which gender is naturally adept for learning English.	0.585	
5. Please indicate which gender performs well in English classes .	0.630	
6. Please indicate which gender is generally good at learning English according to your English instructors.		0.694
7. Please indicate which gender is eager and motivated to learn English according to your English instructors.		0.716
8. Please indicate which gender is good at using the methods and strategies that are effective for learning English according to your English instructors.		0.619
9. Please indicate which gender is naturally adept for learning English according to your English instructors.		0.560
10. Please indicate which gender performs well in English classes according to your English instructors.		0.717

Table 4.15 presents the fit indices and comparative fit indices of the revised version of the QLGB along with two other models (one-factor model and the higher order model). The results showed that the proposed two factor model and the higher order model provided a good model fit to the data (CFA and TLI ≥ 0.95). Further scrutiny demonstrated that the two-factor model had lower AIC and aBIC values which indicated that this model was relatively better than the higher order model. The inter-factor correlation in this model was statistically significant ($r = .58, p < .001$).

The standardised factor loadings were also analysed to evaluate the adequacy of the proposed model. As seen in Table 4.14, all items statistically significantly loaded on their hypothesised factors with factor loadings ranging between .49 and .71. Overall, the results supported that the items corresponded to and are a strong predictor of their priori specified factors.

Table 4.15. Goodness of fit indices of the QLGB - the Main Study

Model & Number of factors	χ^2	<i>df</i>	RMSEA	CFI	TLI	SRMR	AIC	aBIC
1. One-factor model	237.992** *	30	.107	.786	.678	.076	10565 .18	1060 8.47
2. Proposed two-factor model with correlated residual variance	46.210*	29	.031	.982	.972	.037	10271 .30	1031 5.84
3. Alternate higher order model	44.616*	45	.031	.983	.972	.037	10273 .30	1031 9.07

* $p < .05$. ** $p < .01$. *** $p < .001$.

Internal Consistency Reliability. As seen in Table 4.16, Cronbach's alphas were above the cut-off point (i.e. $> .70$) for both sub-scales of the QLGB. The values of skewness and kurtosis indicated that the data was not normally distributed.

Table 4.16. Scale Statistics of the QLGB – The Main Study

	No. of Items	Possible Range	Observed Range	M	SD	Skewness	Kurtosis	Cronbach's α
Learner Beliefs	5	5-25	3-25	13.77	2.48	-.221	4.757	.75
Teacher Beliefs	5	5-25	5-25	14.18	2.27	-.187	4.188	.79

Gender Stereotype Endorsement

To establish whether the participants' views were stereotypical (i.e., whether the participants endorsed the commonly accepted gender stereotyped beliefs in L2 learning), a series of one-sample *t*-tests were performed on the belief scores in SPSS v24. The analyses were done for sub-samples of males and females independently. Basically, a one sample *t*-test informs us about the extent to which participants' responses are away from an identified neutral point. In the current study, as the possible range of the scores obtained from the sub-sections of the QLGB ranged between 5-25 (5 = females are always better; 25 = males are always better), a neutral point of 15 (females and males are equally better) was identified. As such, the analyses aimed to determine the extent to which the participants' scores about gender differences about L2 learning (i.e., their beliefs about L2 learning and perceptions about their teachers' beliefs) were significantly divergent from the neutral point of 15. That is, a gender stereotyped belief that females are better at L2 learning was expected to be statistically significantly away from the midpoint of 15 and towards the 25 end of the scale. A significant *t*-test *p*-value would mean that the sample endorsed the gender stereotypes. If the *p*-value is not significant, this means the scores do not vary far enough away from 15 and therefore that the participants' beliefs are not distinguished as stereotypical (they are just beliefs about competence rather than stereotypical beliefs).

First, a one sample *t*-test was run on the learner beliefs. Consistent with the expectations, the gender stereotype that females are better at L2 learning was significantly endorsed by both females ($t(302) = -9.96, p < .001$) and males ($t(307) = -7.58, p < .001$). With regards to the participants' beliefs about their teachers' perceptions of L2 learning, females endorsed that their teachers believed more females than males were better at L2 learning ($t(301) = -8.46, p < .001$). Similarly, males also believed that their teachers

had gender stereotypical beliefs about L2 learning ($t(305) = -4.28, p < .001$). The results showed that the responses significantly varied from the neutral point of 15. That is, the participants endorsed the commonly accepted gender stereotyped beliefs.

4.4.3.2. The Multidimensional Language Class Anxiety Scale

CFA. To confirm the results obtained from the pilot study data, a CFA was performed on the main study data to verify the factor structure, dimensionality, and internal consistency of each of the five sub-scales of the MLCAS. All analyses were run using maximum-likelihood estimation in *Mplus* 7.4 (Muthen & Muthen, 2013). Unlike the pilot study, 2% of the data was missing in the main study. This was handled using full information maximum likelihood (FIML) in *Mplus* 7.4 (Muthén & Muthén, 2013). According to Arbuckle (1996) and Enders and Bandalos (2001), FIML is more efficient to deal with missing data than the traditional methods such as listwise and pairwise deletion. FIML computes parameter estimates in the model on the basis of all available data, including the incomplete cases. It has a number of theoretical and empirical advantages such as reducing bias resulting from missing data and preserving the overall power of the analysis (Graham et al., 2012).

Additionally, based on the theory and logic behind L2 learning discussed in the literature review, five plausible alternative models were hypothesised which could represent the structure of the MLCAS. A second CFA was run to test these models. The structures of the model with the best fit and of the others are explained in detail below:

Hypothesised models of L2 anxiety.

Model 1: Lower order correlated residuals. Model 1 hypothesises that thirty items form into five lower order factors namely *Listening Activity Anxiety (LAA)*, *Reading Activity Anxiety (RAA)*, *Writing Activity Anxiety (WAA)*, *Speaking Activity Anxiety (SAA)* and *Classroom Testing Anxiety (CTA)*. It is also assumed that the lower order factors correlate with each other. Six items are attributed to each L2 domain. Within each L2 domain, there were two cognitive items, two affective items, two physiological items. In this model, the residuals variances of cognitive, affective and physiological items across different L2 domains were allowed to correlate (Figure 4.2).

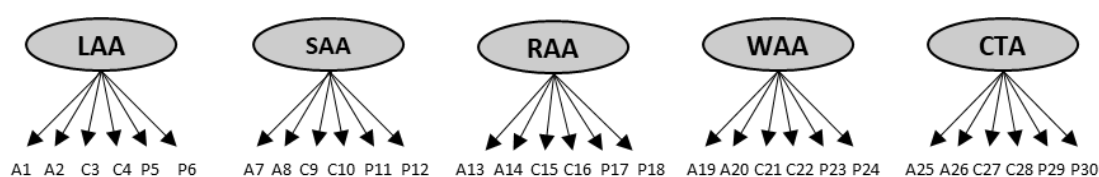


Figure 4.2. *Lower Order Correlated Residuals. For simplicity, the relations between corresponding residuals were omitted.*

Model 2: Higher order correlated residuals. This model is an extension of Model 1. In addition to five lower order factors, in this model, it is hypothesised that all covariance between lower order factors can be explained by a single higher order factor which is termed *Language Class Anxiety (LCA)*. Correlated residual variances of cognitive, affective and physiological items across L2 domains were also included in this model (Figure 4.3).

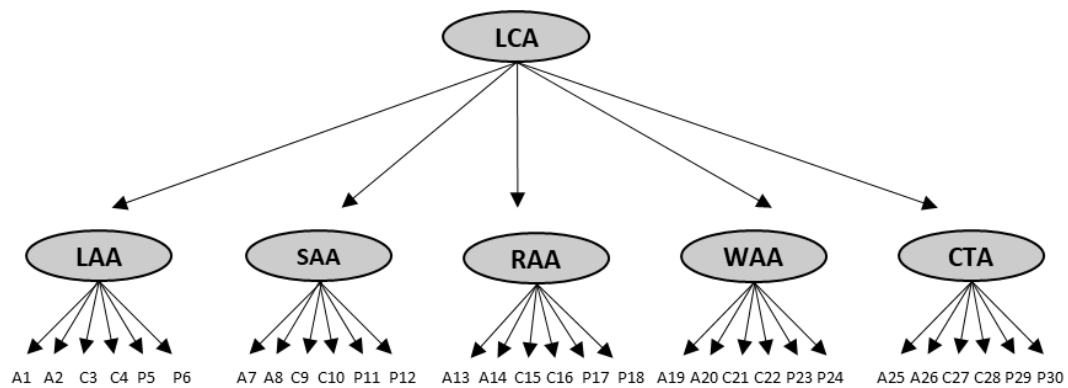


Figure 4.3. *Higher Order Correlated Residuals. For simplicity, the relations between corresponding residuals were omitted.*

Model 3: Lower order L2 anxiety with three method factors. In Model 3, instead of correlating the residuals of cognitive, affective and physiological items across L2 domains, these are introduced to the model as three method factors. It is hypothesised that the items load on both their theoretical constructs which are SAA, LAA, WAA, RAA and CTA and the latent method factors which are cognitive, affective and physiological constructs (Figure 4.4).

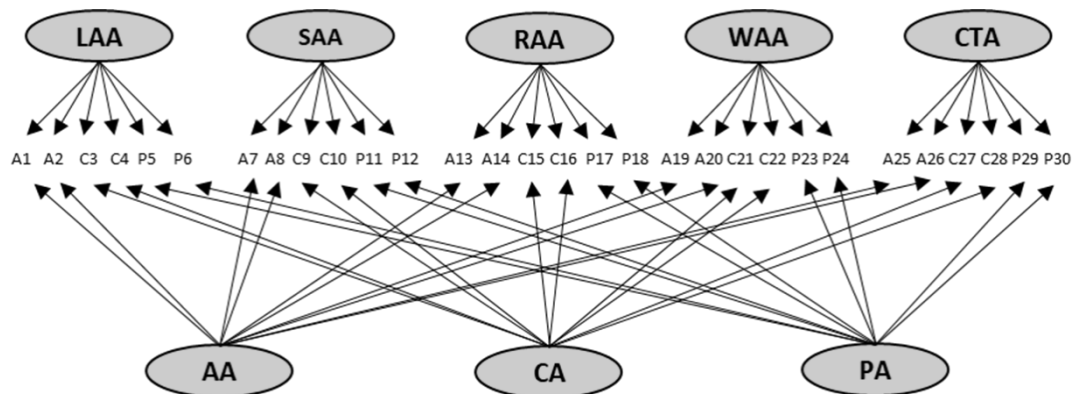


Figure 4.4. *Lower Order L2 Anxiety with Three Method Factors.*

Model 4: Bifactor model with correlated residual variance. In Model 4, all the items load on one general factor representing *Language Class Anxiety*. Additionally, there are five group factors (SAA, LAA, WAA, RAA and CTA) which represent common factors measured by the items and potentially explain item response variance which could not be accounted for by the general factor. Correlated residual variances of cognitive, affective and physiological items across L2 domains were also included and allowed to correlate with each other. The general and group factors were uncorrelated (Figure 4.5).

Model 5: Bifactor model without correlated residual variance. Model 5 is almost the same as Model 4 except that in Model 5, correlated residual variances of cognitive, affective and physiological items across L2 domains were not included.

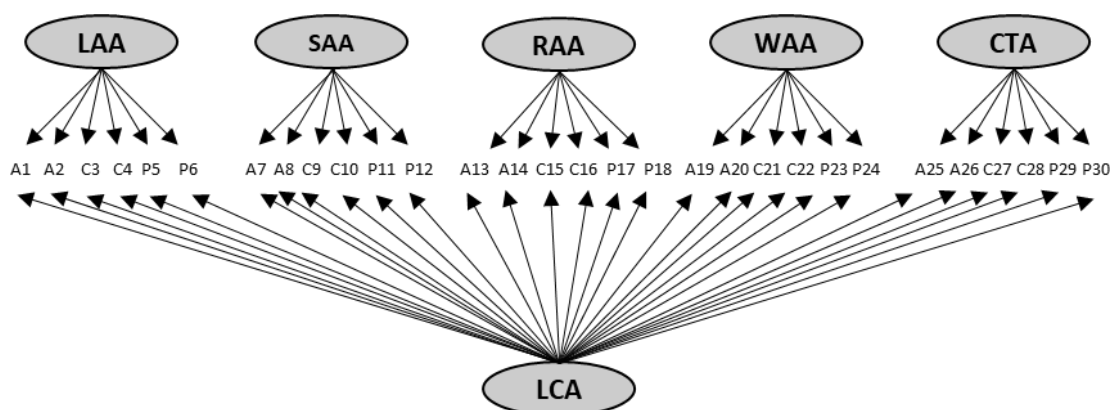


Figure 4.5. *Bifactor model with correlated residual variance.* For simplicity, the relations between corresponding residuals were omitted.

The same model-fit criteria which has been described above were used to evaluate the acceptability of all analysed models. The goodness of fit indices for all subscales are given in Table 4.17. The CFA results indicated that all the subscales of the MLCAS displayed adequate fit to the main study data. Table 4.18 presents the standardised factor loadings for the subscales of the MLCAS. Factor loading estimates showed that all the items loaded onto their hypothesised factors (range of R^2 s = .61–.91).

The results from the second CFA are also provided in Table 4.19. The CFA results showed that Model 4, the bifactor model with correlated residual variance, provided a better fit to the data compared to the other models. This shows that the MLCAS consists of five group factors namely LAA, WAA, SAA, RAA and CTA and one general factor namely L2 anxiety.

Table 4.17. Standardised factor loadings for the subscales of the MLCAS – The Main Study

Items	Factors				
	RAA	WAA	LAA	SAA	CTA
1. Thinking that I will be doing a reading activity in class makes me feel uneasy.	0.810				
2. I feel nervous during reading activities in class.	0.648				
3. Even before reading activities in class, I worry whether I will be able to understand the material.	0.792				
4. I worry whether I am sufficiently prepared for reading activities in class.	0.773				
5. I get tense during reading activities in class (e.g., Having tense muscles).	0.819				
6. When I don't understand something important during reading activities in class, my heart is beating fast.	0.640				
7. Thinking that I will be doing a writing activity in class makes me feel uneasy.		0.832			
8. I feel nervous during writing activities in class.		0.766			
9. Even before writing activities in class, I worry whether I will be able to understand the material.		0.620			
10. I worry whether I am sufficiently prepared for writing activities in class.		0.813			
11. I get tense during writing activities in class (e.g., Having tense muscles).		0.824			
12. When I don't understand something important during writing activities in class, my heart is beating fast.		0.700			
13. Thinking that I will be doing a listening activity in class makes me feel uneasy.			0.667		
14. I feel nervous during listening activities in class.			0.886		
15. Even before listening activities in class, I worry whether I will be able to understand the material.			0.676		

16. I worry whether I am sufficiently prepared for listening activities in class.	0.814
17. I get tense during listening activities in class (e.g., Having tense muscles).	0.800
18. When I don't understand something important during listening activities in class, my heart is beating fast.	0.708
19. Thinking that I will be doing a speaking activity in the class makes me feel uneasy.	0.835
20. I feel nervous during speaking activities in class.	0.840
21. Even before speaking activities in class, I worry whether I will be able to understand the material.	0.693
22. I worry whether I am sufficiently prepared for speaking activities in class	0.613
23. I get tense during speaking activities in class (e.g., Having tense muscles).	0.838
24. When I don't understand something important during speaking activities in class, my heart is beating fast.	0.637
25. Thinking that I will be taking an English language test makes me feel uneasy.	0.774
26. I feel nervous while taking an English language test.	0.918
27. Even before taking an English language test, I worry whether I will be able to understand the material.	0.611
28. I worry whether I am sufficiently prepared for taking an English language test.	0.641
29. I get tense while taking an English language test (E.g. Having tense muscles).	0.831
30. When I don't understand something important while taking English language test, my heart is beating fast.	0.718

Note: Listening Activity Anxiety (LAA), Reading Activity Anxiety (RAA), Writing Activity Anxiety (WAA), Speaking Activity Anxiety (SAA), Classroom Testing Anxiety (CTA)

Table 4.18. The goodness of fit indices for the variables in the MLCAS – The Main Study

	χ^2	df	RMSEA	CFI	TLI	SRMR	AIC	Abic
SAA	15.354*	6	.052	.993	.982	.017	9058.77	9083.77
LAA	20.222**	6	.064	.986	.964	.018	8638.16	8663.16
WAA	10.361	6	.035	.996	.991	.012	8506.22	8531.21
LAA	5.411	5	.012	1.000	.999	.012	8678.55	8704.74
CTA	26.946***	5	.087	.984	.952	.023	9147.19	9173.37

* $p < .05$. ** $p < .01$. *** $p < .001$.

Note: Listening Activity Anxiety (LAA), Reading Activity Anxiety (RAA), Writing Activity Anxiety (WAA), Speaking Activity Anxiety (SAA), Classroom Testing Anxiety (CTA)

Table 4.19. The goodness of fit indices for all hypothesised models for the MLCAS - Main Study

Model	χ^2	df	RMSEA	CFI	TLI	SRMR	AIC	aBIC
1	776.00***	261	.003	.946	.911	.077	41766.31	42046.40
2	798.737***	266	.001	.945	.909	.075	41787.76	42061.87
3	2137.71***	375	.090	.815	.786	.343	42978.68	43121.50
4	482.180***	241	.042	.975	.954	.030	41156.92	41459.21
5	1231.13***	375	.063	.910	.896	.059	41841.94	41984.75

* $p < .05$. ** $p < .01$. *** $p < .001$.

Factor loading estimates obtained from the domain factors revealed that there were four items which had very low factor loadings ($\lambda=0.02-0.16$) (see Table 4.20). These items were 'When I don't understand something important during reading activities in class, my heart races' (Item 6) and its equivalences for LAA and SAA (Item 18 and 24) and 'I get tense during listening activities in class (e.g., Having tense muscles)' (Item 17). However, all the other items loaded onto their hypothesised factors (range of λ s = .20 - .70). The low factor loadings are not unusual in bifactor models where the majority of variance is accounted for by the general factor. As hypothesised, the four items with low factor loadings highly loaded onto the general factor ($\lambda = .65 - .85$) as well as the other 26 items ($\lambda = .49 - .80$).

Table 4.20. Standardised factor loadings for the bifactor model with correlated residual variance – The Main Study

Items	Factors					
	RAA	WAA	LAA	SAA	CTA	LLA
1. Thinking that I will be doing a reading activity in class makes me feel uneasy.	0.562					0.621
2. I feel nervous during reading activities in class.	0.376					0.528
3. Even before reading activities in class, I worry whether I will be able to understand the material.	0.284					0.718
4. I worry whether I am sufficiently prepared for reading activities in class.	0.410					0.651
5. I get tense during reading activities in class (e.g., Having tense muscles).	0.484					0.733
6. When I don't understand something important during reading activities in class, my heart is beating fast.	0.051					0.656
7. Thinking that I will be doing a writing activity in class makes me feel uneasy.		0.595				0.588
8. I feel nervous during writing activities in class.		0.633				0.490
9. Even before writing activities in class, I worry whether I will be able to understand the material.		0.301				0.550
10. I worry whether I am sufficiently prepared for writing activities in class.		0.499				0.622
11. I get tense during writing activities in class (e.g., Having tense muscles).		0.459				0.715
12. When I don't understand something important during writing activities in class, my heart is beating fast.		0.201				0.804
13. Thinking that I will be doing a listening activity in class makes me feel uneasy.			0.640			0.523
14. I feel nervous during listening activities in class.			0.332			0.757
15. Even before listening activities in class, I worry whether I will be able to understand the material.			0.563			0.588

16. I worry whether I am sufficiently prepared for listening activities in class.	0.351	0.716
17. I get tense during listening activities in class (e.g., Having tense muscles).	0.169	0.795
18. When I don't understand something important during listening activities in class, my heart is beating fast.	-0.054	0.854
19. Thinking that I will be doing a speaking activity in the class makes me feel uneasy.	0.700	0.589
20. I feel nervous during speaking activities in class.	0.688	0.604
21. Even before speaking activities in class, I worry whether I will be able to understand the material.	0.317	0.609
22. I worry whether I am sufficiently prepared for speaking activities in class	0.321	0.514
23. I get tense during speaking activities in class (e.g., Having tense muscles).	0.578	0.537
24. When I don't understand something important during speaking activities in class, my heart is beating fast.	0.020	0.828
25. Thinking that I will be taking an English language test makes me feel uneasy.	0.578	0.580
26. I feel nervous while taking an English language test.	0.527	0.699
27. Even before taking an English language test, I worry whether I will be able to understand the material.	0.247	0.622
28. I worry whether I am sufficiently prepared for taking an English language test.	0.463	0.540
29. I get tense while taking an English language test (E.g. Having tense muscles).	0.423	0.711
30. When I don't understand something important while taking English language test, my heart is beating fast.	0.239	0.741

Note: Listening Activity Anxiety (LAA), Reading Activity Anxiety (RAA), Writing Activity Anxiety (WAA), Speaking Activity Anxiety (SAA), Classroom Testing Anxiety (CTA) and Language Class Anxiety (LCA)

Overall, the results from the main study confirmed that the MLCAS does not only consist of a single common factor that represents L2 anxiety which is the multidimensional construct, but also addressed the five individual factors which are LAA, WAA, RAA, SAA and CTA that compromise it. It also accounts for the variance due to the different components of anxiety which are cognitive, physiological and affective.

Assessment of Internal Consistency. The descriptive statistics including observed ranges, means, standard deviations, skewness, kurtosis and internal consistency (Cronbach α) of the subscales are shown in Table 4.21. As seen, all the sub-scales showed good internal consistency (Cronbach's α ranging from .88 to .89). As the values for both skewness and kurtosis were within ± 1 , it can be concluded all the variables were normally distributed.

Table 4.21. Item and Scale Statistics for Model 4 – The Main Study

	No. of Items	Possible Range	Observed Range	M	SD	Skewness	Kurtosis	Cronbach Alpha
SAA	6	6-30	6-30	16.42	5.59	.109	-.613	.88
LAA	6	6-30	6-30	14.97	5.55	.261	-.266	.89
RAA	6	6-30	6-30	13.60	5.03	.559	-.301	.88
WAA	6	6-30	6-30	14.21	5.29	.495	-.007	.89
CTA	6	6-30	6-30	17.85	5.90	-.22	-.617	.89
LCA	30	30-150	30-147	77.05	23.58	.081	-.099	.96

Note: Listening Activity Anxiety (LAA), Reading Activity Anxiety (RAA), Writing Activity Anxiety (WAA), Speaking Activity Anxiety (SAA), Classroom Testing Anxiety (CTA) and Language Class Anxiety (LCA)

Predictive validity of the MLCAS. To establish the predictive validity of the MLCAS, it was also explored whether there is a link between the students' overall language performance scores and the MLCAS results. A latent correlation analysis was run in *Mplus* using the main study sample data to examine correlations between students' overall language performance scores and the subscales of the MLCAS. The

performance scores were measured by the language tests which students took after completing their one-year English language programme. The tests included all the four language skills namely reading, writing, listening and speaking. The results showed that all the variables were significantly and negatively correlated with the performance scores (see Table 4.22). This affirmed the predictive validity of this new questionnaire.

Table 4.22. The domains of the MLCAS - Correlation analysis

		Performance
1.	SAA	-.189***
2.	RAA	-.298***
3.	WAA	-.266***
4.	LAA	-.202***
5.	CTA	-.243***
6.	LCA	-.269***

* $p < .05$. ** $p < .01$. *** $p < .001$.

Note: Listening Activity Anxiety (LAA), Reading Activity Anxiety (RAA), Writing Activity Anxiety (WAA), Speaking Activity Anxiety (SAA), Classroom Testing Anxiety (CTA) and Language Class Anxiety (LCA)

4.4.3.3. The Questionnaire of Self-Efficacy Beliefs in Learning a New Language (QSLL)

CFA. The 11 items retained through the EFA and reliability analysis were modelled within a CFA using the main study data. In addition to the two-factor solution, correlated residual variance was introduced for each language skill which was informed by the theory of L2 teaching and learning. The CFA was run using maximum-likelihood estimation and full information maximum likelihood (FIML) to deal with missing data (J. W. Graham et al., 2012). The factor structure was assessed using a number of goodness of fit indices which were the Root Mean Square Error of Approximation (RMSEA), Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), the Standardised Root Mean Square Residual (SRMR), Akaike Information Criterion (AIC) and sample-size adjusted Bayesian Information criterion (aBIC). Hu and Bentler (1999) state that an acceptable model is indicated by $RMSEA < .05$, $SRMR < .08$, and CFI and $TLI > .95$. As for AIC and aBIC, the model with the smallest value is recommended (Hu & Bentler, 1999).

Table 4.23 presents the fit indices and comparative fit indices of the revised version of the QSLL along with three other models (one-factor model with and without correlated

residual variance and an alternate two-factor model without correlated residual variance). The two-factor model (Figure 4.6) with correlated residual variance showed a good fit to the data which were superior to the other three models. The CFA results supported the two-factor solution recommended by the EFA. As seen, the model was able to achieve the acceptable level of fit to the data (CFA and TLI ≥ 0.90). The inter-factor correlation was statistically significant ($r = .73$, $p < .001$).

Table 4.23. Goodness of fit indices of the QSL - the Main Study

Model & Number of factors	χ^2	<i>df</i>	RMS EA	CFI	TLI	SRM R	AIC	aBIC
1. One-factor model with correlated residual variance	161.011***	34	.080	.937	.898	.054	1534 6.95	15397. 27
2. One factor model without correlated residual variance	406.45***	44	.119	.820	.775	.076	1564 0.57	15679. 84
3. Proposed two-factor model with correlated residual variance	69.895***	33	.044	.982	.970	.031	1524 4.46	15296. 83
4. Alternate two- factor model without correlated residual variance	143.453***	43	.063	.950	.936	.045	1531 5.36	15355. 83

* $p < .05$. ** $p < .01$. *** $p < .001$.

The adequacy of this model was also determined in relation to the standardised factor loadings which are indicated in Table 4.24. All items statistically significantly loaded on their hypothesised factors with factor loadings ranging between .63 and .80. This confirmed that the items corresponded to and a strong predictor of their priori specified factors.

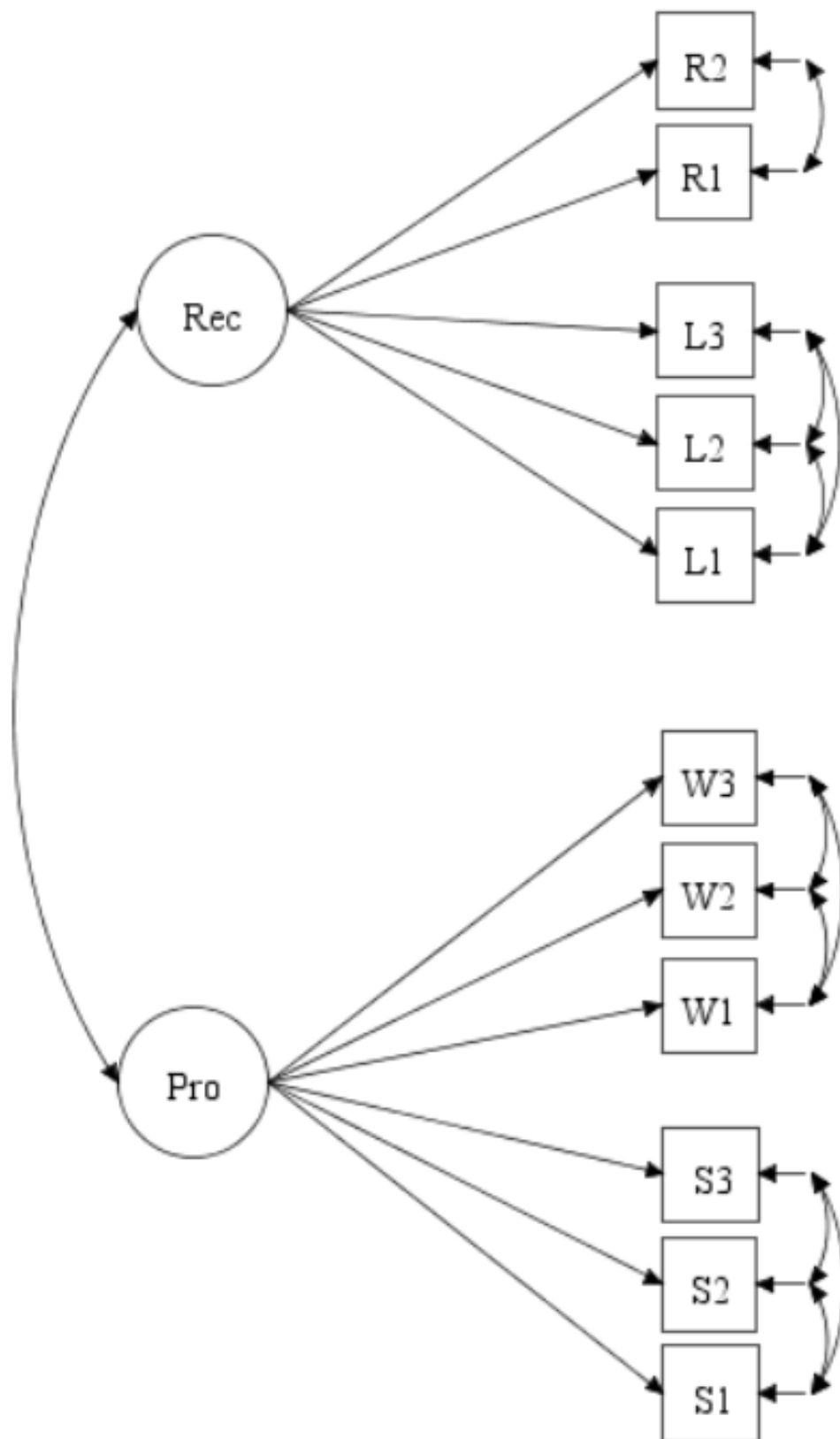


Figure 4.6. *Two-factor model for QSL*

Table 4.24. Standardised Loadings for the Two-factor Solution – The Main Study

Items	Reception	Production
1. I can listen to and understand the main point of an English radio/TV program on a personal /professional interest.	0.679	
2. I can watch and understand English TV news programs without English/Turkish subtitles.	0.668	
3. I can watch and understand English films and TV series without English/Turkish subtitles.	0.674	
4. I can read and understand the main point of English articles and reports concerned with contemporary problems without using any kind of dictionaries	0.802	
5. I can read and understand the majority of long and complex English literary texts such as novels and essays without using any kind of dictionaries.	0.717	
6. I can have a conversation with my classmates and instructors on familiar and daily topics such as families, hobbies, work and travel in English without any preparation in advance.		0.577
7. During the English class, I can ask questions to my instructors and answer their questions verbally in English.		0.634
8. I can verbally state my opinions about the contemporary issues or my plans for the future in English.		0.638
9. I can write a personal letter/an email describing my experiences and impressions in English without using any kind of dictionaries.		0.756
10. I can write an English essay giving reasons in support of or against a particular point of view without using any kind of dictionaries.		0.662
11. I can express myself in clear well-structured written English text, expressing points of view at some length without using any kind of dictionaries.		0.726

Internal Consistency Reliability. As seen in Table 4.25, Cronbach's alphas were $\geq .84$ for both sub-scales of the QSLL. Also, all variables were normally distributed as shown by the skewness and kurtosis statistics.

Table 4.25. Scale Statistics of the QSLL – The Main Study

	No. of Items	Possible Range	Observed Range	M	SD	Skewness	Kurtosis	Cronbach's α
Reception	5	5-25	3-25	13.71	4.08	.200	.019	.85
Production	6	6-30	3-30	20.24	4.35	-.247	.339	.84

Predictive validity of the QSLL. To establish the predictive validity of the QSLL, the link between the students' overall language performance scores and the QSLL results was also explored. A latent correlation analysis was performed in *Mplus* using the main study sample data to examine correlations between students' language performance scores and the two subscales of the QSLL. The performance scores were measured by the language tests which students took after completing their one-year English language programme. The tests included all the four language skills namely reading, writing, listening and speaking. The results showed that both reception and production were significantly and positively correlated with the performance scores ($r_s = .207$ and $.447$; $p < .001$, respectively). This affirmed the predictive validity of this new questionnaire.

4.5. Structural Equation Modelling (SEM) Results

All the structural equation models were built and estimated in *Mplus* 7.4 using the MLR estimator (maximum likelihood with robust standard errors) which adjusts standard errors and chi-square goodness of fit statistics to enhance robustness against the non-normal distribution. To evaluate each model, a number of model fit indices which are provided by *Mplus* were utilised. Some of the model fit indices which guided our analysis are chi-square (χ^2), Root Mean Square Error of Approximation (RMSEA), the Standardised Root Mean Square Residual (SRMR), Comparative Fit Index (CFI), and the Tucker Lewis Index (TLI). According to Hu and Bentler (1999), general threshold levels for an acceptable model are $RMSEA \leq 0.05$, $SRMR \leq 0.08$, and CFI and $TLI \geq 0.95$.

To examine how L2 self-efficacy and L2 anxiety mediate relations between learners' perceptions about their teachers' gender stereotyped beliefs about foreign language learning and their overall performance, the model which is depicted in Figure 4.7 was

built and tested using the structural equation modelling (SEM). Because the items in the Gender Stereotyped Beliefs Questionnaire was rated using a “1-Always Females” and “5-Always Males” scale, higher scores would mean beliefs in favour of males (i.e., males are more successful in L2 learning than females) and lower scores would mean beliefs in favour of females (i.e., females are more successful in L2 learning than males).

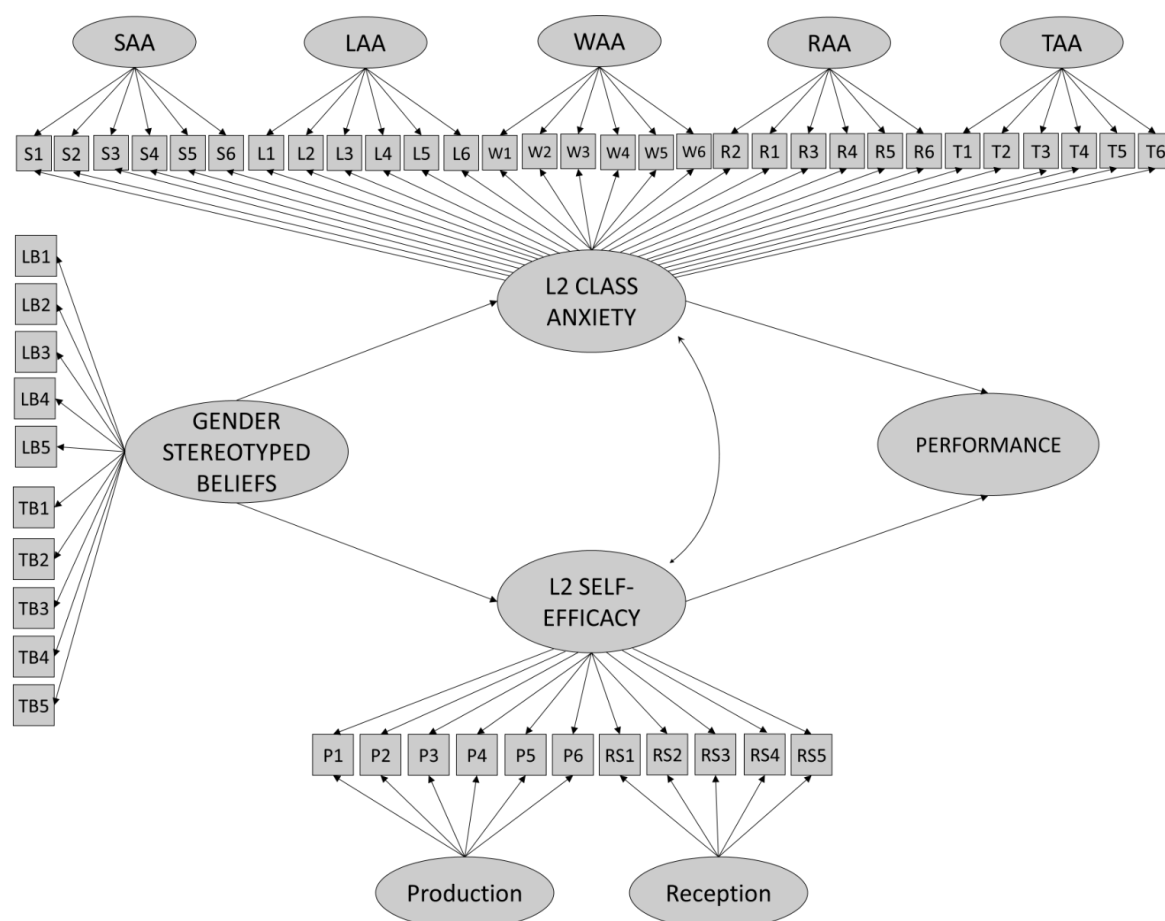


Figure 4.7. *The SEM Model*

4.5.1. The Relationship between Learners' Gender Stereotyped Beliefs about Language Learning and their Language Performance

4.5.1.1. Gender as a Moderator

Based on L2 learning theory and practice, it was hypothesised that gender moderates the relationship between gender stereotyped beliefs and the mediators, L2 anxiety and self-efficacy (Figure 4.8). In moderation, it is tested whether the regression of a dependent variable (i.e., L2 anxiety and self-efficacy) on an independent variable (i.e., gender stereotyped beliefs) depends on the level of a third variable (i.e., gender). A moderator variable affects the strength and direction of the relationship between the dependent and independent variables.

Moderating Hypothesis: It was basically expected that the paths from learners' gender stereotyped beliefs to L2 anxiety and L2 self-efficacy would be different across genders. While strong gender stereotype endorsement (i.e., females are better than females) would mean higher L2 anxiety and lower L2 self-efficacy for males, for females, it would mean higher L2 self-efficacy and lower L2 anxiety.

Alternative Hypothesis: The alternative hypothesis for the current study was that gender did not play a moderating role in the link between the learners' gender stereotyped beliefs and the mediators, L2 anxiety and L2 self-efficacy. That is, it was expected that the paths coefficients did not significantly differ between males and females.

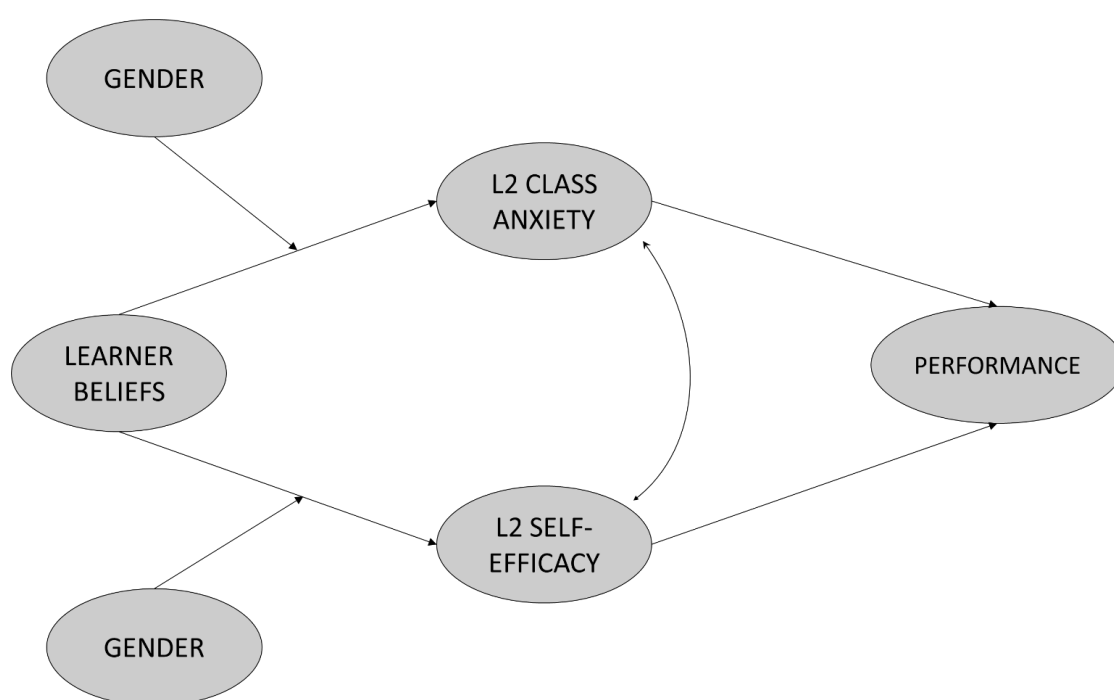


Figure 4.8. *The moderation model for learners' gender stereotyped beliefs*

To determine whether to accept the moderating hypothesis, a series of multi-group analyses were conducted using the GROUPING command on *Mplus*. To statistically test the moderation by gender, two models were created. The first model was the baseline model where the parameters were freely estimated across the groups. In the second model, the two paths from gender stereotyped beliefs to L2 Anxiety and to L2 self-efficacy were constrained to be invariant between groups. The models were compared using the Satorra-Bentler scaled chi-square difference test (Satorra & Bentler, 1994).

When there is a significant difference between the chi-square values of the models, it can be concluded that the null hypothesis (i.e., that gender is not a moderator) is rejected.

When the chi-square values of the two models were compared, a statistically significant difference was detected between the baseline model ($\chi^2(1803) = 2561.132$, $p < .001$) and the constrained model ($\chi^2(1805) = 2570.252$, $p < .001$) (Table 4.26). The results confirmed that gender moderated the relationship between the gender stereotyped beliefs and one or both mediators. To establish which path was moderated by gender, the paths were sequentially unconstrained. The Chi-square difference results are provided in Table 4.27 and 4.28 below.

Table 4.26. The Chi-square difference test between the baseline model and the constrained model

Model	Chi-square	Degrees of Freedom	P-value	Gender as a Moderator
Unconstrained model	2561.132*	1803		
Constrained model	2570.252*	1805		
Chi-Square Difference test	8.64	2	0.013	Yes

Table 4.27. The Chi-square difference test between the baseline model and the model for L2 anxiety

Model	Chi-square	Degrees of Freedom	P-value	Gender as a Moderator
Unconstrained model	2561.132*	1803		
Constrained model 1 (L2 Anxiety)	2563.058*	1804		
Chi-Square Difference test	1.92	1	.165	No

Table 4.28. The Chi-square difference test between the baseline model and the model for L2 Self-efficacy

Model	Chi-square	Degrees of Freedom	P-value	Gender as a Moderator
Unconstrained model	2656.623*	1803		
Constrained model 2 (L2 self-efficacy)	2570.460*	1804		
Chi-Square Difference test	9.32	1	.002	Yes

The chi-square difference test results revealed that while gender moderated the relationship between gender stereotyped beliefs and L2 self-efficacy ($\chi^2(1) = 9.32, p = .002$), it did not serve as a moderator of the relationship between gender stereotyped beliefs and L2 anxiety ($\chi^2(1) = 1.92, p = .165$). That is, although the standardised estimates were different across the groups, they did not significantly differ between males and females. Thus, it can be concluded that the null hypothesis is partially rejected in the current model. Therefore, the SEM analyses were conducted separately for males and females. Standardised coefficients were calculated separately for females and males using the GROUPING command in *Mplus*.

4.5.1.2. Path Analysis

The hypothesised SEM model is presented in Figure 4.9. The total sample size for this analysis was 694 (males = 343; females = 351). The missing data patterns ($n = 84$; males = 46, females = 38) were handled using the FIML. The model provided a good fit to the data: $\chi^2(1803) = .2561.132$, RMSEA < .03, SRMR = .054, CFI = .946, TLI = .936.

The results for the path analyses are provided in Figure 4.10. Among females, higher gender stereotyped beliefs (i.e., that they are a male domain) were associated with lower L2 self-efficacy ($\beta = -.211, p = .014$), but they were not significantly related to L2 anxiety ($\beta = .055, p = .510$). As for language performance, a greater L2 self-efficacy was strongly associated with higher language performance ($\beta = .399, p < .001$). However, L2 anxiety was not associated with language performance as such ($\beta = -.071, p = .422$). There was a statistically significant negative correlation between L2 anxiety and L2 self-efficacy ($\beta = -.601, p < .001$).

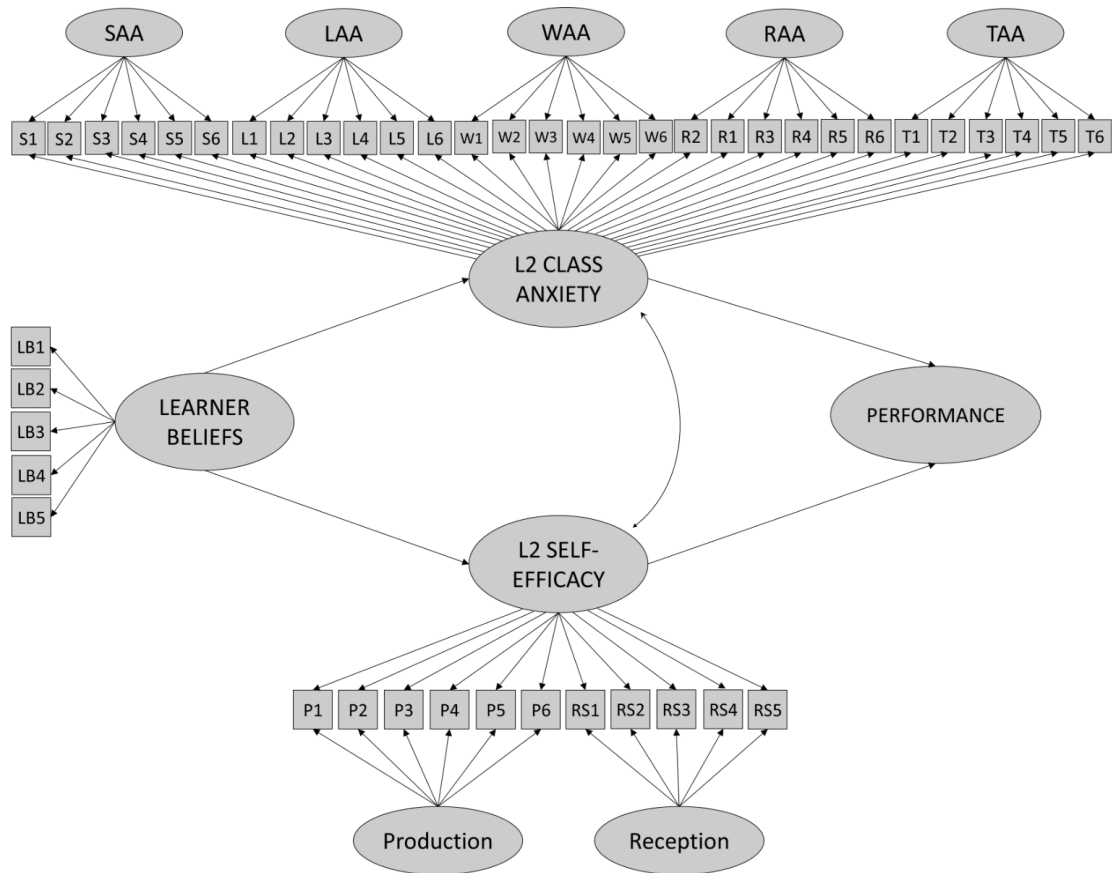


Figure 4.9. *The SEM Model for Learners' Gender Stereotyped Beliefs*

Using the MODEL INDIRECT and CINTERVAL commands in *Mplus*, the indirect paths from gender stereotyped beliefs to language performance and confidence intervals (CI) were examined. The significance of the indirect effects depends on both the lower limit and the upper limit values of 95% CIs. When 95% CIs do not cross zero, indirect paths are statistically significant at $p < .05$. That is, the lower and upper limit values of 95% CIs must be both positive or negative (MacKinnon, 2012).

A significant indirect path was observed between gender stereotyped beliefs and language performance through L2 self-efficacy beliefs ($\beta = -.084$, $SE = .036$, 95%CIs $[-.025, -.143]$). Those who held a belief that L2 learning was a female domain had higher performance scores due to higher L2 self-efficacy. The path from gender stereotyped beliefs to language performance was not statistically significant when mediated by L2 anxiety ($\beta = -.004$, $SE = .008$, 95%CIs $[-.017, .009]$). The total indirect path from gender stereotyped beliefs to language performance was statistically significant ($\beta = -.088$, $SE = .038$, 95%CIs $[-.026, -.150]$).

The path from gender stereotyped beliefs to L2 self-efficacy was statistically significant for males ($\beta = .141$, $p = .044$). However, L2 anxiety was not related to gender stereotyped

beliefs ($\beta = -.098$, $p = .194$). The relationship between L2 anxiety and language performance was not statistically significant ($\beta = -.059$, $p = .528$). An increased L2 self-efficacy was associated with higher language performance ($\beta = .386$, $p < .001$). As is the case with females, L2 self-efficacy was negatively correlated L2 anxiety ($r = -.685$, $p < .001$).

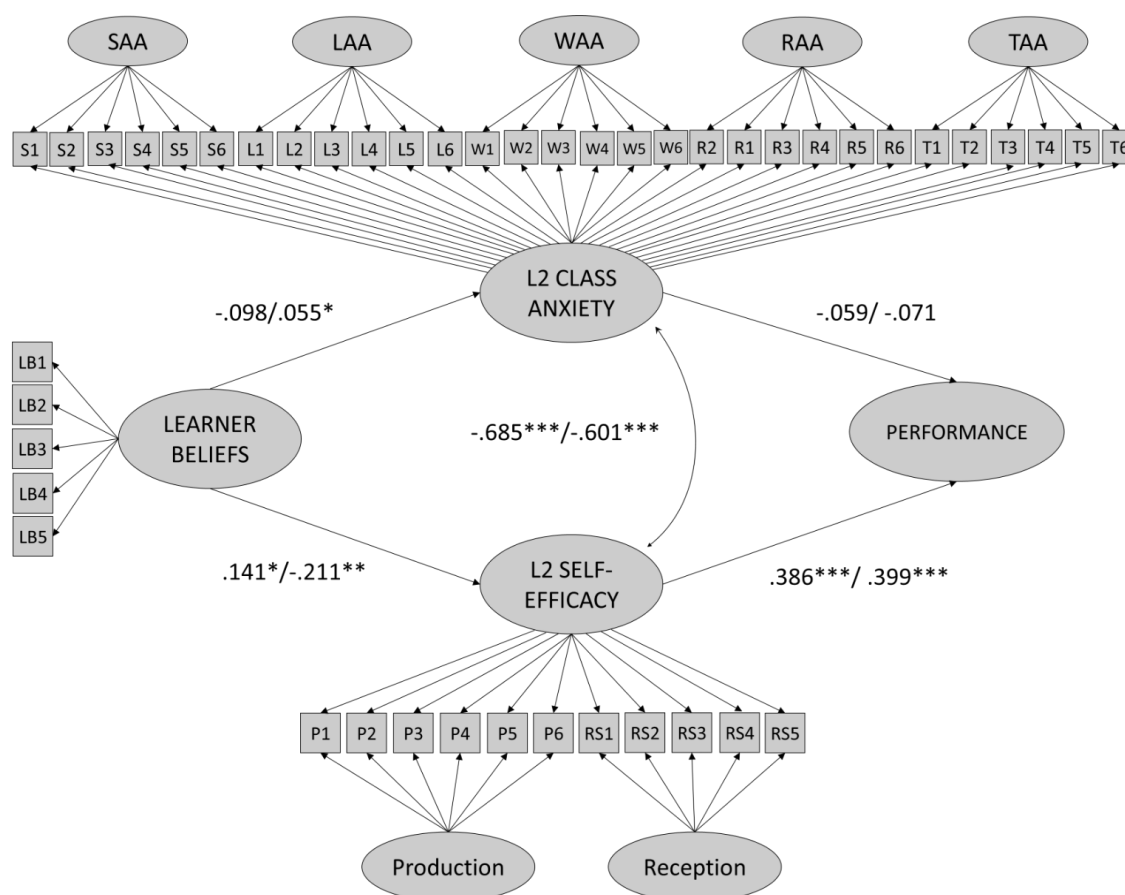


Figure 4.10. Standardised path coefficients of the hypothesised model for males (before the dash) and females (after the dash) (* $p < .05$, ** $p < .01$, *** $p < .001$).

The results concerning the indirect paths from gender stereotyped beliefs to language performance demonstrated that neither L2 anxiety ($\beta = .006$, $SE = .011$, 95%CI $[-.012, .024]$) nor L2 self-efficacy ($\beta = .055$, $SE = .033$, 95%CI $[-.001, .108]$) had an indirect effect on males' language performance. The total indirect effect from gender stereotyped beliefs to language performance was relatively significant for males ($\beta = .060$, $SE = .032$, 95%CI $[-.113, .008]$).

4.5.2. The Relationship between Learners' Perceptions about their Language Teachers' Gender Stereotyped Beliefs and Language Performance

4.5.2.1. Gender as a Moderator

In addition to the relationship between learners' gender stereotyped perceptions about L2 learning and the mediators, L2 anxiety and L2 self-efficacy, the link between learners' perceptions about their L2 teachers' gender stereotyped beliefs and the mediators was also investigated. It was tested to determine whether gender played a moderating role in the link between the learners' perceptions of their teachers' beliefs and their L2 anxiety and L2 self-efficacy. The proposed SEM model is presented in Figure 4.11.

Moderating Hypothesis: It was expected that the paths from learners' perceptions of their teachers' beliefs to L2 anxiety and L2 self-efficacy would be different across genders. It was hypothesised that when males strongly believed that their L2 teachers found females more successful than males, they would have higher L2 anxiety and lower L2 self-efficacy for males. In contrast, females who held the same belief would have higher L2 self-efficacy and lower L2 anxiety.

Alternative Hypothesis: The alternative hypothesis for the current study was that gender did not play a moderating role in the link between the learners' perceptions of their L2 teachers' beliefs and the mediators, L2 anxiety and L2 self-efficacy. That is, it was expected that the paths coefficients did not significantly differ between males and females.

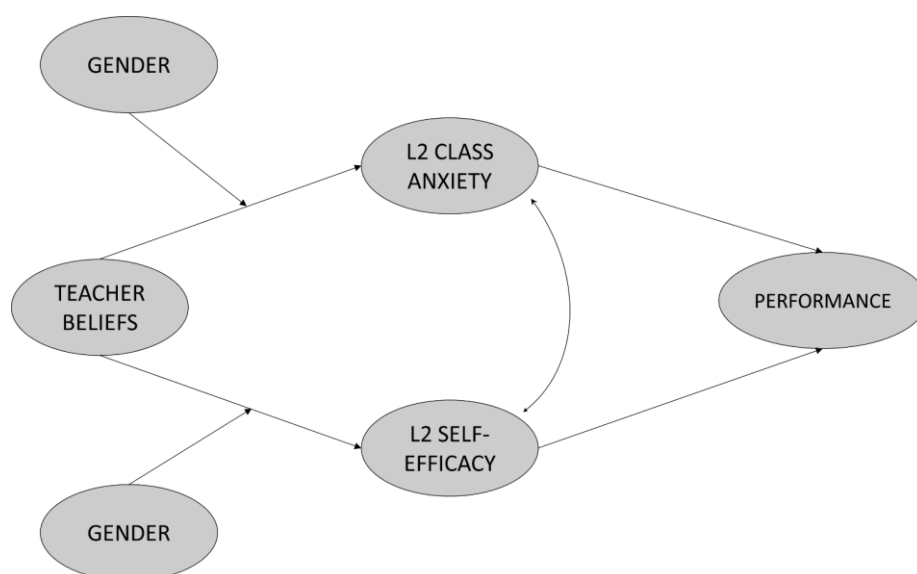


Figure 4.11. *The Moderation Model*

To check these hypothesis, multigroup analyses were conducted using the GROUPING command on *Mplus*. The first step was to estimate the paths in the original model for females and males simultaneously. The resulting model was referred to as the baseline model (or free or unconstrained model). In this model, the estimates of the paths were allowed to differ across the groups. Following this, a constrained model was created. In this model, the parameter estimates of the paths from gender stereotyped beliefs to L2 anxiety and L2 self-efficacy were constrained to be equal across the groups. The comparison between the models was made based on the Satorra-Bentler scaled chi-square difference test (Satorra & Bentler, 1994). According to this test, if the chi-square values are significantly different across the baseline and constrained models, the alternative hypothesis (i.e., that gender is a moderator) cannot be rejected.

The results for the baseline model ($\chi^2(1803) = 2656.623$, $p < .001$) indicates that the model fits the data for both females and males. The chi-square difference tests for gender as a moderator is provided in Table 4.29 below. The chi square (χ^2) difference test shows that there is a significant difference between the baseline model ($\chi^2(1803) = 2656.623$, $p < .001$) and the constrained model ($\chi^2(1805) = 2674.448$, $p = .001$) which means that the path estimates for females and males are different in at least one path. Thus, it can be concluded that the moderating hypothesis is accepted.

To identify the specific path that differ significantly between females and males, the paths were sequentially unconstrained in the constrained model. This helped identify the specific path that differ significantly between females and males. The chi-square significance test results are provided in Table 4.30 and 4.31 below.

Table 4.29. The Chi-square difference test between the baseline model and the constrained model

Model	Chi-square	Degrees of Freedom	P-value	Gender as a Moderator
Unconstrained model	2656.623*	1803		
Constrained model	2674.448*	1805		
Chi-Square Difference test	17.825	2	.001	Yes

The results confirm that the models for the paths from gender stereotyped beliefs to L2 anxiety ($\chi^2(1803) = 2671.136$, $p < .001$) and L2 self-efficacy ($\chi^2(1803) = 2669.412$, $p = .001$) are significantly different from the baseline model. As the null hypothesis was rejected by the results, it was decided that SEM analyses were conducted separately for males and females. Standardised coefficients were estimated for males and females separately using the GROUPING command in *Mplus*.

Table 4.30. The Chi-square difference test between the baseline model and the model for L2 anxiety

Model	Chi-square	Degrees of Freedom	P-value	Gender as a Moderator
Unconstrained model	2656.623*	1803		
Constrained model 1 (L2 Anxiety)	2671.136*	1804		
Chi-Square Difference test	14.513	1	.001	Yes

Table 4.31. The Chi-square difference test between the baseline model and the model for L2 Self-efficacy

Model	Chi-square	Degrees of Freedom	P-value	Gender as a Moderator
Unconstrained model	2656.623*	1803		
Constrained model 2 (L2 self-efficacy)	2669.412*	1804		
Chi-Square Difference test	11.18	1	.001	Yes

4.5.2.2. Path analysis

The hypothesised SEM model is presented in Figure 4.12 below. The total sample size for the SEM analysis was 694 (males = 343; females = 351) which means a total of 84 data patterns (males = 44; females = 40) was missing from the data. This was handled using full information maximum likelihood (FIML) in *Mplus* 7.4 (Muthén & Muthén, 2012). Based on the aforementioned criteria, the model yielded an acceptable fit to the data:

$\chi^2(1803) = .2656.623$, $RMSEA < .037$, $SRMR = .055$, $CFI = .940$, $TLI = .928$. The results for the SEM analyses are presented in detail below.

Figure 4.13 presents the standardised path coefficients of the hypothesised model for males and females. For females, an increase in gender stereotyped beliefs (i.e., that they are a male domain) was associated with higher L2 anxiety ($\beta = .255$, $p = .001$) and lower L2 self-efficacy ($\beta = -.287$, $p = .001$). L2 self-efficacy was negatively correlated L2 anxiety ($r = -.565$, $p < .001$). L2 anxiety was not significantly related to language performance ($\beta = -.069$, $p = .437$). However, a greater L2 self-efficacy was associated with higher language performance ($\beta = .399$, $p < .001$).

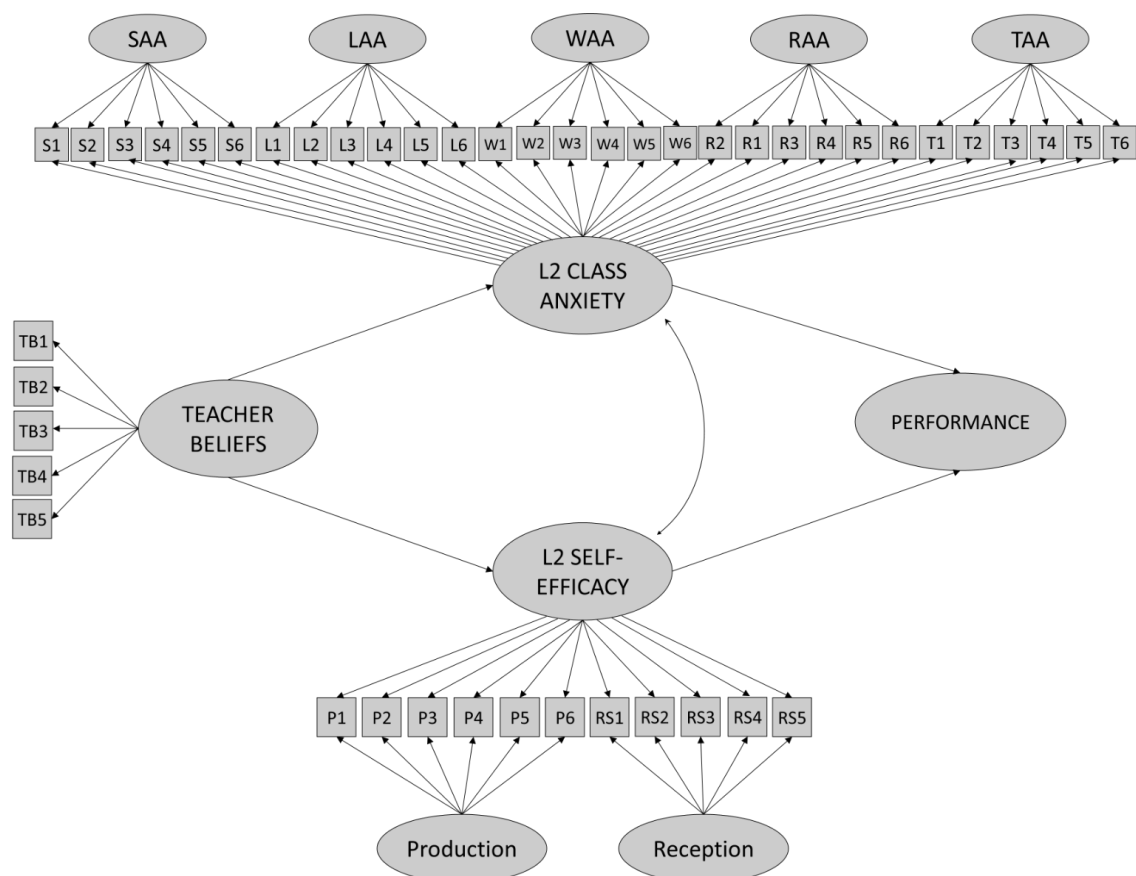


Figure 4.12. *The SEM Model for Students' perceptions about their Language Teachers' Gender Stereotyped Beliefs*

The indirect paths from gender stereotyped beliefs to language performance were also estimated. The results showed that the indirect path from gender stereotyped beliefs to performance through L2 self-efficacy was significant ($\beta = -.114$, $SE = .040$, 95%CIs $[-.049, -.179]$). This means that participants who perceived their teachers to hold a belief that L2 learning was a female domain performed better due to higher L2 self-efficacy.

The other indirect path which was from gender stereotyped beliefs to language performance through L2 anxiety was not statistically significant for females ($\beta = -.018$, $SE = .024$, 95%CIs $[-.056, .021]$) because the CI included zero. L2 anxiety did not serve as a mediator between gender stereotyped beliefs and language performance. The total indirect path from gender stereotyped beliefs to language performance was statistically significant ($\beta = -.132$, $SE = .040$, 95%CIs $[-.066, -.198]$).

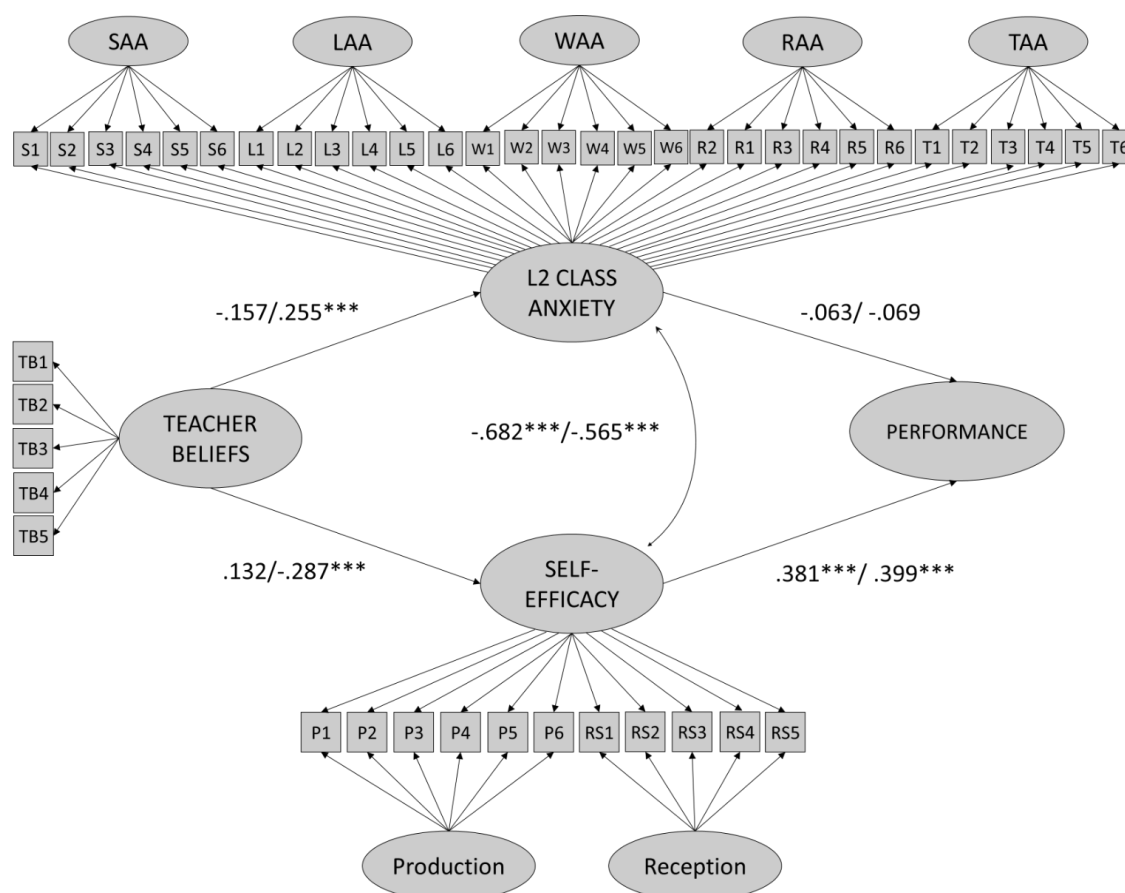


Figure 4.13. Standardised path coefficients of the hypothesised model for males

(before the dash) and females (after the dash) (* $p < .05$, ** $p < .01$, *** $p < .001$).

For males, the path from gender stereotyped beliefs to L2 anxiety ($\beta = -.157$, $p = .083$) and L2 self-efficacy ($\beta = .132$, $p = .114$) was not statistically significant. As for the paths to language performance, higher L2 self-efficacy beliefs were associated with higher language performance ($\beta = .381$, $p < .001$). However, L2 anxiety was not related to language performance ($\beta = -.063$, $p = .492$). Similar to females, L2 self-efficacy was negatively correlated with L2 anxiety ($r = -.682$, $p < .001$).

The indirect effects from gender stereotyped beliefs to language performance were not statistically significant for males. L2 anxiety ($\beta = .010$, $SE = .016$, 95%CIs $[-.016, .036]$)

and L2 self-efficacy ($\beta = .050$, $SE = .035$, 95%CI [-.007, .108]) were not a mediator of the relationship between gender stereotyped beliefs and language performance. The total indirect path from gender stereotyped beliefs to performance was not statistically significant ($\beta = .06$, $SE = .037$, 95%CI [-.001, .122]).

4.6. Discussion

The goal of this study was to determine whether L2 anxiety and L2 self-efficacy mediated the relationship between language learners' gender stereotyped beliefs and their L2 performance. To answer the research questions, a demographically diverse group of Turkish learners studying EFL at university level was recruited. Findings revealed that the extent to which learners' level of stereotype endorsement affects their L2 performance through the mediating role of L2 anxiety and L2 self-efficacy depends on their gender. By conducting a moderated mediation analysis, it was found that gender moderated the mediating effects of L2 anxiety and L2 self-efficacy. Therefore, the data was analysed for females and males separately.

First, learners' gender stereotyped beliefs and how these were linked to their performance through L2 anxiety and L2 self-efficacy was investigated. The results showed that L2 anxiety was not a mediator of the relationship between gender stereotyped beliefs and L2 performance for either group. L2 self-efficacy mediated the relationship between the gender stereotyped beliefs and L2 performance only for females. Females who held the belief that L2 learning is a female domain showed greater self-efficacy and higher L2 performance. L2 anxiety was negatively correlated with L2 self-efficacy for females and males.

In addition to the learners' own gender stereotyped beliefs, their perceptions about their teachers' gender stereotyped beliefs were also examined. The results were similar to the findings above. Confirming the aforementioned findings, it was found that L2 anxiety did not mediate the link between the learners' perceptions of their teachers' gender stereotyped beliefs neither females nor males. As for L2 self-efficacy, it was a mediator only for females. Females who believed that their teachers thought language learning was a female domain were found to have more self-efficacy and higher L2 performance. L2 self-efficacy was negatively correlated with L2 anxiety for both females and males.

Based on the findings, it can be concluded that gender stereotyped beliefs do not affect L2 learners' performance through their L2 anxiety. The beliefs are mostly associated with

L2 learners' self-efficacy, but only for females. Males' L2 self-efficacy do not seem to be influenced by any kind of gender stereotyped beliefs tested. These results suggested that self-efficacy is at least one important mechanism by which gender stereotyped beliefs are related to females' L2 performance. It is important to note that the non-significant results might be due to the shared variance between L2 anxiety and L2 self-efficacy. That is, it is highly likely that due to the large amount of shared variance between L2 self-efficacy and L2 anxiety, there was not enough unique variance left for L2 anxiety to predict the participants' performance. Therefore, precaution should be taken in interpreting the results.

The findings are both congruent and contradicting with the existing literature. The results concerning L2 self-efficacy suggest that there are differences between females and males in terms of the link between the gender stereotyped beliefs and L2 performance. While the belief that females are better at L2 learning than males was found to be linked to higher self-efficacy for females, this belief did not lead to significant decrease in males' self-efficacy. This finding is consistent with the notion that socialisation processes might play a role in the formation of the self-beliefs (Skaalvik & Skaalvik, 2004). As discussed in Chapter 1 and 2, the differences between females and males are clear-cut in Turkey. While females are believed to be weak and fragile, males are always expected to be strong and confident. Even though such differences are not necessarily related to L2 learning, it is possible that these are reflected on L2 learning experience in a number of ways. For example, it is highly likely that Turkish female L2 learners are more susceptible to gender stereotyped beliefs than their male peers due to the gender roles and expectations of females and males in the Turkish society. It can be claimed that women who experience inequality (i.e., being secondary to men) everyday might become more sensitive to any kind of gender stereotypes and receive their effects more easily. Also, the aforementioned status of females and males in Turkey might have an impact on males as well. It is highly likely that males were less willing to communicate negative self-judgements. They might even tend to overestimate their own performance in order to maintain their dominant position. These findings corroborate the ideas of Mills (2014) who suggested that more research is needed to better understand the relationship between gender and self-efficacy beliefs in L2 learning given the role of socialisation process in the formation of self-related beliefs. This issue is followed up in the following study.

Even though most researchers have found that there is a negative correlation between language anxiety and achievement (e.g., Aida, 1994; Horwitz, 1986; Saito et al., 1999),

evidence presented in this study suggests that L2 anxiety is not the mechanism by which gender stereotypes are associated with a negative performance outcome. As this is a surprising finding, some consideration should be given why this might be the case for the L2 learners involved in the current research. As stated above, it seems possible that such finding is due to shared variance between L2 self-efficacy and L2 anxiety. In addition to this, another likely explanation might be that of Attentional Control Theory proposed by Eysenk et al. (2007). According to this theory, anxiety causes processing inefficiency by deteriorating the central executive functions. However, this inefficiency does not result in poor performance all the time because anxious individuals adopt compensatory strategies such as putting extra effort to achieve success. Considering this with regard to L2 learners who participated in this study, there might be some external factors which played a role in determining the effect of gender stereotypes on their language performance through L2 anxiety. For these L2 learners, it was compulsory to attend the preparatory classes and pass the final English proficiency test so that they could start studying their chosen subjects (e.g., engineering, economics, education etc.) at university. Poor English performance would mean that they might have to study English for another year, which would result in losing both time and money for the L2 learners and their families. Instead of experiencing such difficulties, the L2 learners might have put extra effort (i.e., compensatory strategy) which might have reduced the debilitating effects of gender stereotyped beliefs via L2 anxiety.

Also, as evidenced by the one sample *t*-tests (please see Section 4.4.3.1 for the one-sample *t*-tests), English was not as gendered as some of the academic domains tested in the current research. Therefore, it is possible that the belief that English is a female domain was not strong enough among the participants, which might have resulted in the non-significant link between the gender stereotyped beliefs and the performance via L2 anxiety and L2 self-efficacy (for males only).

The non-significant link between L2 performance and L2 anxiety may also be explained by the fact that levels of L2 anxiety might vary depending on different cultural groups (Horwitz, 2001). For example, as noted by Horwitz (2001), when Horwitz (1986) and Aida (1994) examined American foreign language learners' L2 anxiety using the FLCAS, they found similar levels of L2 anxiety among two different groups. Compared to these studies, Truitt (1996) who collected data from Korean EFL learners using the FLCAS found relatively higher levels of L2 anxiety. In contrast, Kunt (1999) indicated that the levels of L2 anxiety in Turkish and Turkish-Cypriot learners of English were somewhat lower than the previous studies. It is also asserted that L2 anxiety might vary depending

on classroom environment. For example, in a class where the teacher is perceived as a threat, L2 anxiety might be high. However, if the teacher is perceived as a source of support and language learners have a strong relationship with their teacher, it is highly likely that their L2 anxiety is lower (Horwitz, 2001). Given the varying levels of L2 anxiety among different language learners depending on L2 learning situation, the rather contradictory result gained in this study might suggest that L2 anxiety was not in fact a significant issue among the chosen sample of Turkish adults learning EFL at university level.

One final point about the rather contradictory result gained in Study 1 may be due to the way L2 anxiety and performance was measured in the current study. I used the overall L2 anxiety and performance scores rather than focusing on anxiety and performance concerned with each language skill (i.e., listening, speaking, reading and writing) discretely. As discussed in Chapter 2 at length, each language skill is distinct from each other, and they have their own characteristics and dynamics. It is highly likely that L2 learners' level of anxiety in each skill and related performance might differ from each other due to differences in their strengths and weaknesses (Horwitz, 2001).

4.6.1. Limitations and Suggestions for Future Research

The findings in this study are subject to at least two limitations. First, gender is treated as a binary construct in the current study. As previous research has mostly focused on differences between males and females, there is little research looking at the dynamics within the groups of males and females including this research. However, as Eckert and McConnell-Ginet (1992) state, women and men are not pre-set groups. There are considerable differences between males and females and also, within each gender group, among females and among males. These differences are the result of a variety of different social practices within the communities alongside different personal attributes and power relations. These practices construct the notion of gender in various, subtle and changing ways (Eckert & McConnell-Ginet, 1992). As such, although there are similar patterns when it comes to gender stereotyping and its effects, the results cannot be generalised for all males or females. That is, group membership (being a male/female) may mean different things to different people. According to Social Identity Theory, people do not automatically have a sense of belonging in the groups that surround them (Hogg, 2006). The idea of groupness occurs when people identify themselves with the groups. Therefore, the extent to which the gender stereotypes and gender stereotyping of academic subjects have an effect on a person's self-efficacy beliefs in chosen academic domains may depend on their group identification level.

Future research should therefore concentrate on the investigation of differences within as well as across two gender groups.

Secondly, the conclusions drawn in this research were based on the self-report instruments apart from the performance measure, so any implications should be interpreted taking this into consideration. The self-report instruments may fall short when it comes to assessing the participants' feelings and understandings. Participants might prefer sharing their "ideal self" rather than "actual self". Even though they are assured that the information given is confidential and not shared with the third parties, it is still almost impossible to detect in self-report questionnaires whether participants fill in the questionnaires appropriately (Duckworth & Yeager, 2015) . Therefore, it is safe to say that the findings presented here are provisional and should be treated cautiously until more research has been conducted to replicate the results. Also, adopting a different method of inquiry such as classroom observation would help establish a greater degree of accuracy on the issue of gender stereotyping of L2 learning. As such, it is suggested that future research investigates the phenomenon of gender stereotyping of L2 learning applying other methods of inquiry.

4.6.2. Original Contribution to Knowledge

Despite the aforementioned limitations, this research has three contributions to literature and educational practice. First of all, even though gender stereotyping of academic domains and its effects on academic achievement has been a popular area of research, the substantial body of research focuses on the impact of gender stereotyping on academic subjects such as maths and science. However, there is considerable future research potential in researching other academic subjects such as foreign language learning which can further our understanding of gender stereotyping of academic subjects and their impact on academic achievement. This is particularly important in the context chosen in the current study as the stakes are so high. The empirical research presented in this study represents the first to examine university level EFL learners' gender stereotyped beliefs pertaining to language learning and their link to their performance via L2 anxiety and L2 self-efficacy. This dynamic nature of the study is a good response to the call for a dynamic approach in L2 teaching and learning research.

Also, it is generally believed that gender stereotyping concerns females and their academic achievement in so-called masculine academic subjects only. However, it is possible that due to having low self-efficacy and high anxiety, male students are also in danger of underperforming in academic subjects which are mostly associated with

females. Therefore, instead of taking the previous studies and assumptions for granted, this study focused on the extent to which the phenomenon of gender stereotyping of academic domains could be an issue for males as well, especially in a context such as Turkey where there are clear-cut gender roles and expectations of males and females.

Lastly, as discussed above, some of the findings were rather contradictory compared to previous studies. Such difference between this study and the previous ones confirms that gender stereotyped beliefs and their effects cannot be generalised for all communities as the phenomenon of gender stereotyping might change from community to community due to different gender socialisation processes. It is also possible that the effects of gender stereotyping of L2 learning manifest themselves through different mechanisms in different contexts. The current study sheds light on two mechanisms, namely L2 anxiety and L2 self-efficacy and offered some important insights about their mediating roles. A natural progression of this work is to continue focusing on males and their potential underachievement in so-called feminine academic subjects by focusing on different mechanisms as well as L2 anxiety and L2 self-efficacy. This will help determine the extent to which gender stereotyping of L2 learning is linked to L2 performance through various mechanisms.

4.7. Chapter Summary

Chapter 4 has presented the first empirical study which examined the extent to which Turkish adult EFL learners endorsed the common gender stereotypes pertaining to L2 learning and whether there was a relationship between these gender stereotyped beliefs and L2 learners' L2 performance via L2 anxiety and L2 self-efficacy. The chapter has commenced with a report on the development and validation of the three instruments (i.e., the MLCAS, the QSLL, and the QLGB) used to collect data in Study 1. This was followed by the examination of the data gained through these three questionnaires. First, it has been found that both L2 females and males endorsed the common gender stereotype that females were better at L2 learning than males. Following this, the link between L2 learners' gender stereotyped beliefs (both their own beliefs and perceptions about their teachers' beliefs) and their L2 performance via L2 anxiety and L2 self-efficacy has been examined. The findings have indicated that neither L2 self-efficacy nor L2 anxiety played a mediating role for L2 male learners. As for L2 female learners, the link was only mediated via L2 self-efficacy. The chapter has been concluded with the discussion of the findings along with the limitations of this study and suggestions for further research. The results of this study are highlighted further in Chapter 7 which

integrates the findings of all the three studies and discuss them with references to the research literature and research questions framing this thesis.

CHAPTER 5: STUDY 2

5.0. Chapter Outline

Chapter 5 focuses on the second empirical study which qualitatively investigated language learners' and teachers' perceptions about these gender stereotypes and the extent to which they believed such stereotypes related to L2 achievement. The study further aimed to determine whether learners and teachers had a mutual understanding of the root of gender stereotypical perceptions, particularly in relation to what the socialisation processes and agents are in Turkey. More specifically, the role of L2 teachers, as agents of socialisation, in the gender issues in L2 learning was explored. The chapter consists of three main parts. The first part reviews the literature on the existing gender stereotypes in L2 learning and identifies the current gaps in the literature. It further outlines the approaches taken in the current study to address these research gaps. The second part describes the qualitative design of Study 2 in detail and presents the findings gained from the semi-structured interviews. Finally, the last part discusses the limitations of the study and the implications for further research.

5.1. Introduction

In many L2 learning environments, there is a commonly held belief that L2 learning is a feminine domain (Carr & Pauwels, 2006; Pomerantz, 2008; Schmenk, 2004). In their study, for example, Carr and Pauwels (2006) interviewed 200 boys aged between 12 and 18 in the major English-dominant communities in Australia (i.e., Queensland, the Australian Capital Territory and Victoria) and examined their perceptions about FL learning. The majority of the boys indicated that L2 learning is not something 'real boys' do or are good at, and due to this belief, they do not prefer studying languages beyond compulsory level. Similarly, Plante, Theoret and Favreau (2009) conducted a quantitative study among 1137 Canadian French speaking students who were recruited from Grade 6, 8 and 10. Consistent with the previous literature, this study also revealed that L2 learning was strongly associated with females by the students in all grades and both genders.

Despite gender stereotypes pertaining to L2 learning being a feminine domain being consistent across research studies, it is important to note that gender and how it is perceived by people change across cultures, time and different communities of practice (Schmenk, 2004). Therefore, the current qualitative research aims to question the assumed stereotypes and explore some of their complexity. As seen in the examples of Plante et al. (2009) and Carr and Pauwels (2006), several studies investigating gender

stereotyping of FL learning have concentrated on primary or secondary school students in the English-speaking countries. As such, little is known about the extent to which the phenomenon in question is applicable to other groups of learners in other contexts such as adults or young adults studying a new language at university level or beyond. In an attempt to address this limitation, Pomerantz (2008) conducted a qualitative discourse-analytic study examining two male learners learning Spanish at advanced level at a US university. The data showed that even if both participants embraced the normative gender identities, they positioned themselves differently in the language classrooms. One of the learners preferred conforming to the general dominant ideology (i.e., males are not good language learners) by not participating in the activities and receiving low grades in the exams. In contrast, the other learner challenged this ideology by positioning himself as a good language learner who was active during the classes and received higher grades in the exams. However, he also aligned himself well with the other male students by behaving in certain ways such as making jokes which was mostly associated with males. Such findings suggest that both femininity and masculinity might be linked to L2 achievement, but this link is not established in similar ways in all L2 learning contexts. Considering there is little empirical data available as to whether the commonly accepted gender stereotypes in L2 learning exist in the non-English speaking countries as well, the current study aims to extend the literature by investigating this phenomenon among adult L2 learners learning English as a foreign language (EFL) at university level in Turkey.

The possible sources of the existing gender stereotypes have also been investigated in the literature. Some researchers have argued that the gender stereotypic approaches to certain academic subjects might stem from the differences in ability. However, according to van der Vleuten (2016), ability does not fully explain such approaches because there are several studies showing that females' and males' performance did not fit the expected patterns. For example, it has been demonstrated that females outperformed males in the academic subjects which are mostly associated with males (e.g., STEM) (see Stoet & Geary, 2018). Similarly, in a recent study, Wucherer and Reiterer (2016) analysed the data from 64 participants (32 females) and concluded that depending on the type of language task, males' language performance was better than females. While males performed better in the tasks related to pronunciation, females achieved a more desirable outcome in the grammar-related tasks. As such, van der Vleuten (2016) suggests that gender stereotypes are created and reinforced not due to the gender differences in ability per se, but through the gender socialisation process in which it is almost always possible to find some degree of gender role stereotyping. According to the gender socialisation theories (Fagot et al., 2000), individuals are attributed with

certain characteristics of femininity and masculinity. Those who prefer to conform to gender role expectations imposed socially and culturally behave in certain ways in order not to experience the uncertainty that standing out might create (van der Vleuten et al., 2016). As in some societies, it is traditionally believed that males are breadwinners and women are the ones who provide caring and nurturing, males and females choose their career paths accordingly, which results in gender imbalances in certain academic subjects. As a result, academic subjects are labelled as feminine or masculine. According to Piller and Pavlenko (2007), many academic subjects concerned with multilingualism (e.g., language teaching, translating or interpreting) are highly feminised. It is further asserted that such areas are relatively poorly paid jobs with little job security (Piller & Pavlenko, 2007), which does not align well with males' 'breadwinner' role. As such, males might have less tendency to choose such academic paths compared to females.

According to Bussey and Bandura (1999), there are some agents who play a crucial role in the aforementioned gender socialisation process. Among these come family, peers, media, and school (Bussey & Bandura, 1999). It is believed that after the family, the school is the second major agent of gender socialisation because students spend the majority of their time at school with teachers, so they are influential role models (Stromquist, 2007). According to Stromquist (2007), many teachers state that they do not treat female and male students differently. However, it may not always be the case in practice (Stromquist, 2007). Although it is not always visible, teachers might be biased towards males and females in the classroom due to their gender stereotyped beliefs. In one of the studies, for example, Tiedemann (2002) analysed the effect of teachers' gender stereotypes on their impressions of their students' competence and effort in maths. The results showed that teachers believed that male students had greater maths ability than female students. Also, that male students were more capable of logical thought than females, and that maths was a more challenging subject for females than for males. According to Gunderson et al. (2012), being influenced by their teachers' perceptions, students might also start developing negative attitudes towards certain subjects which could potentially influence their performance, subject choice, and career paths. Although there are studies highlighting the link between teachers' gender stereotypes and their students' perceptions about their ability in academic subjects such as maths or science, there have been few investigations into the role of teacher beliefs and behaviours in L2 learning. The main aim of the current research is, therefore, to reveal whether Turkish L2 teachers themselves have any gender stereotyped beliefs about female and male L2 learners and whether they tend to behave differently towards

each gender in the classroom. To provide a full account of the phenomenon, both L2 teachers' and L2 learners' perceptions are explored.

5.2. Research Aims and Questions of Study 2

Differently from Study 1 and Study 3 (see Chapter 6 for Study 3), which adopted a quantitative research approach, Study 2 was qualitative in nature and was designed to provide an in-depth account of L2 learners' and teachers' perceptions of the gender stereotypes pertaining to L2 learning. As it also aimed to help inform the potential sources of any existing gender stereotypes and underpin the results gained in Study 1 and 3, a special focus was put on the role of L2 teachers as an agent of socialisation which the gender stereotype literature does not tend to focus in so much depth on.

The main research questions of Study 2 were as follows:

1. To what extent do Turkish L2 learners and teachers endorse the gender stereotypes pertaining to L2 learning?
2. To what extent does the L2 learning environment in the chosen context foster the salience of gender-stereotyped beliefs?
 - 2.1. Do L2 teachers, as an agent of socialisation, play a role in sustaining and legitimising gender stereotyped beliefs in L2 learning?

5.3. Method

5.3.1. Participants

A total of 32 participants were recruited through purposive sampling. The overall sample was made up of two sub-samples. The first sub-sample constituted 17 Turkish teachers (7 males and 10 females) teaching English as a foreign language (EFL) at university level. The second sub-sample consisted of 15 Turkish EFL learners (8 males and 7 females). As for the EFL teachers, the priority was given to the ones working at the universities which agreed to participate in Study 1. These teachers were approached while collecting data for Study 1 and asked if they would like to volunteer to be interviewed for the current study. The volunteers differed from each other in terms of years of teaching experience. Fifteen participants had between 3 and 7 years of teaching experience. Of the other two participants, one had 9 and the other had 15 years of teaching experience.

Unlike the EFL teacher participants, the EFL learners ($n = 15$) were selected from the universities which were not involved in Study 1. The aim was to increase sample representativeness by reaching the EFL learners who did not have a chance to share

their perceptions in Study 1. At the time of the data collection, the EFL learner participants had been studying English for at least 3 months. There were not any mature learners as all the participants were between 18-21 years of age. The group was diverse in terms of the participants' programme of study. The programmes included, but were not limited to, the following: interior design, architecture, engineering, economics and physics.

5.3.2. Data Collection Procedure

The data was collected using semi-structured interviews. To ensure that the data collection process would run smoothly, the interview questions were piloted with a small group of L2 teachers and learners (2 teachers and 2 learners). The piloting served three main purposes. First, it provided the opportunity to assess the structure (e.g., flow) and the quality (e.g., wording) of the interview questions. Secondly, it enabled the proposed time frame to be tested for feasibility and practicality. Lastly, it gave the chance to practice the interviewing techniques such as building a good rapport with the participants and probing and prompting where necessary. Based on the pilot interviews, some questions were revised for more clarity and conciseness. Also, two questions were omitted to avoid any repetition. The main data collection was conducted using the revised interview schedule. Each interview lasted for 20-30 minutes. The interview questions were based on three broad areas: beliefs about L2 learning, L2 self-efficacy and L2 anxiety. As both L2 teachers and learners were involved in this study, the interview schedule was adjusted for each group accordingly (please see Appendix 5.1 and 5.2 for a copy of each schedule). The details of the interviews are provided in detail below:

Phase 1: Interviews with L2 teachers

The in-depth interviews with L2 teachers aimed to shed light on their perceptions and behaviours towards both male and female L2 learners. The interviews started with the warm-up questions such as "How long have you been teaching English?" and continued with more specific questions concerned with beliefs about L2 learning, L2 self-efficacy and L2 anxiety. In addition to the main questions, some potential probes were also noted to be used when more information needed to be sought. As mentioned above, three issues were initially chosen to be discussed during the interviews. There were as follows:

1. Gender stereotyped beliefs about L2 learning

The purpose of this first issue was to determine whether the participants had any gender stereotypical beliefs about L2 learning. The questions were carefully worded in order not to direct the participants to any particular direction. For example, the questions such as

“Why do you think females are more successful than males” were avoided. Instead, a more neutral way of questioning was preferred (e.g., Have you noticed any differences between male and female students in terms of L2 learning). The open-ended questions aimed to encourage the participants to share their own thoughts, beliefs and experiences. Some of the questions included:

- Do you think there are any gender differences in ability in L2 learning?
- Have you noticed any differences between your female and male students in L2 classrooms?
- Which group is more challenging for you, females or males?
 - a. Why?
 - b. Any examples?

2. L2 self-efficacy

The purpose of the second issue was to gain some in-depth insights about gender stereotyped beliefs pertaining to some specific constructs such as L2 self-efficacy. It was aimed that the findings gained from the interviews could shed more light on the findings of Study 1 and 3 which focused on the link between gender stereotyped beliefs and L2 achievement via the mediating role of L2 self-efficacy. However, the questions were not restricted to L2 self-efficacy. There were also a number of open-ended questions which enabled the participants to offer their original contributions. Some of these questions are provided below:

- Could you please name the most successful three learners in your classes?
- Why do you think these learners are successful?
Prompt: What do these learners have in common?
- What do you think about unsuccessful learners?
Prompt: What do these learners have in common?
 - a. Do you think self-efficacy in language learning affects their language learning performance?
- Have you ever seen any differences between male and female learners in terms of their self-efficacy in language learning?
 - a. Who do you think has lower level of self-efficacy in language learning?
 - b. Why do you think their self-efficacy is low?

3. L2 Anxiety

The third line of questioning asked the participants to share their thoughts, beliefs and experiences on any gender differences pertaining to L2 learners' L2 anxiety. Like in

self-efficacy, this third issue also aimed to inform Study 1 and 3 which investigated the role of L2 anxiety as the mediator of the relationship between gender stereotyped beliefs and L2 achievement.

- Do you think L2 anxiety affects learners' L2 performance?
 - a. How?
- Who do you think tend to be more anxious when it comes to language learning- males or females?
 - a. Why?
 - b. Any examples?

At the end of each interview, the interviewees were asked whether they had any additional thoughts or beliefs they would like to share. This allowed them to offer an original contribution which they considered valuable.

Phase 2: Interviews with L2 learners

The questions for learners intended to explore learners' perceptions about their teachers and their L2 learning experience in general. As in the interviews with the teachers, the interviews with the learners commenced with the warm-up questions such as "How long have you been studying English?" and "Do you like English" and these were followed by some more specific questions pertaining to the three issues discussed above. The learner participants were also asked a number of open-ended questions which enabled them to gain control of the interview process and direct towards the issues which were most important to them. The issues discussed along with the example questions are provided below:

1- Gender Stereotyped beliefs about L2 learning

The purpose of this first line of questioning was twofold. First, it was aimed to explore the learner participants' thoughts, beliefs and experiences in terms of L2 learning and reveal any gender stereotyped beliefs they might hold. Secondly, the participants were asked to share their perceptions about their L2 teachers' thoughts and behaviours. As discussed in the literature review in Chapter 2, even if L2 teachers might believe that they do not hold any gender stereotypes, the impression they leave on L2 learners might be different, and they might play an active role in legitimising and sustaining the common gender stereotypes pertaining to L2 learning. Therefore, the interviews with the learner participants intended to determine the extent to which the learner participants believed that their teachers' thoughts and behaviours were stereotypical and how these related to their L2 learning experiences. Some of the questions used included:

- Do you think there are any gender differences in ability in L2 learning?
Prompt: Are females or males more successful in L2 learning? Why?
- Do you think your teachers hold any gender stereotyped beliefs in terms of L2 learning?
Prompt: Do you think your teachers favour females or males in foreign language classes?
- Do you think your teachers have different attitudes towards females and males in the classroom?
- Do your language teachers have different performance expectations with males and females?

2- *L2 Self-efficacy*

The interviews also aimed to explore the learner participants' perceptions of the role of L2 self-efficacy in L2 learning. They were first asked about whether they believed there were any differences between males and females in terms of L2 self-efficacy. This line of questioning was followed by more specific questions which aimed to enable the participants to reflect on their own L2 self-efficacy and share their beliefs as to how their L2 self-efficacy related to their success in L2 learning. Some of the questions used were as follows:

1. If you compare your male classmates with your female classmates, who do you think have more self-efficacy in L2 learning?
 - a. Why do you think so?
2. What about you? Do you believe that you can learn English easily?
 - a. Why? Why not?

3. *L2 Anxiety*

In addition to L2 self-efficacy, the L2 learner participants were also asked about their thoughts and beliefs about L2 anxiety and how they thought it would relate to their L2 achievement. The purpose of the questions concerned with L2 anxiety was twofold. First, they aimed to investigate the extent to which males and females differed from each other in terms of L2 anxiety. Secondly, the participants were encouraged to share their beliefs and experiences about their own L2 anxiety, if any. Some of the questions used to explore these are presented below:

- If you compare your male friends with your female classmates, who do you think are more anxious during the classes?
 - a. Why?

- b. Any examples?
- What about you? Do you feel always anxious in language classes?
 - a. Why?
 - b. How do you cope with your anxiety?
 - c. Do your teachers help with that?
- How do you think L2 anxiety affects your language learning performance?

The learner participants were also given the opportunity to share their additional thoughts and beliefs at the end of the interviews. The interviews with both the learner and teacher participants were arranged at the convenience of the participants and conducted face to face. At the commencement of each interview, the participants were informed about their right to withdraw as well as being reassured that all the information given would be confidential and it would be reported anonymously. With the permission granted by the participants, all the interviews were digitally recorded in full and transcribed verbatim for analysis.

5.3.2.1. Transcription and Translation

Since previous research has demonstrated that conducting interviews in participants' mother tongue maximizes the quality of data (Twinn, 1998), the interviews in the current study were carried out in Turkish which was the participants' first language. This contributed to the quality of the data in two ways. First, the English language would possibly create a barrier between the L2 learners and me if they were not fluent English speakers at the point of data collection. Second, the L2 teacher participants would not feel comfortable being interviewed in English as they might feel that their English-speaking ability was being judged. Their focus during the interviews might be on their language use rather than the content of their answers. To address such limitations, as a native Turkish speaking researcher, I decided to use Turkish as a means of communication during the interviews.

As the interview questions were originally written in English for the sake of the current thesis, they needed to be translated into Turkish before the interviews. To avoid any lost in meaning across the different versions, the questions were translated using the back-translation method (see Brislin, 1970) which is the most common and highly recommended procedure for cross-language studies (please see 4.3.1 for further information about the back-translation method). As recommended, the questions which were in English were translated in Turkish by two independent translators. The Turkish versions were then back-translated into English. Both Turkish and English versions were evaluated and approved by two different translators.

After the data collection period, the question whether the transcripts were to be translated into English was raised. It is well-recognised in the qualitative research literature that little attention has been given to the translation procedures in cross-language studies (Chen & Boore, 2010; Temple & Young, 2004). Hence, there is not much guidance for researchers as to how to manage these procedures. Within the limited literature, some researchers suggest that where possible, the data should be analysed in the language used in the interviews to protect the authenticity of the data (Chen & Boore, 2010; Temple & Young, 2004). Also, it is noted that translating a high volume of interviews might be time-consuming for researchers (Twinn, 1998). In line with these suggestions and considering the limited time frame and resources available, it was decided that the transcripts were analysed in the source language and the illustrative quotations which would be included in the final report were translated into English utilising the back-translation method.

5.3.2.2. The Role of the Researcher

In qualitative research, it is acknowledged that it is not possible for researchers to free themselves from their own values, beliefs and experiences (Willig, 2013). To establish dependability and trustworthiness of qualitative research, researchers are suggested to maintain reflexivity by reflecting on their role throughout the research process (Willig, 2013). My role as a researcher in the current study was substantially informed by my personal and professional background in Turkey, which held both advantages and disadvantages during the data collection period. First of all, being Turkish and speaking the mother tongue of the participants fluently was an advantage because it enabled me to build a good rapport with the participants easily. Interacting with the participants naturally helped them feel at ease expressing their ideas freely during the interviews. Also, according to Brinkman and Kvale (2015), one of the key virtues of qualitative interviews is empathy. Having lived in the same community as the participants for 25 years and knowing the participants' culture, I could easily grasp the participants' lived experiences in Turkey from the standpoint of an 'insider' and show empathy towards them. Furthermore, I could also listen to the participants from the standpoint of an L2 learner and L2 teacher as I undertook both roles in the past. I started learning EFL at primary school and had been learning it actively until I started studying at university. Having completed my BA in English Language Teaching, I became an EFL teacher and taught EFL at university level for about three years. Having had similar experiences with the participants did not only sustain my understanding of the narratives told by the

participants, but it also helped me ask appropriate follow-up questions giving the participants an opportunity to elaborate their answers.

On the downside, there was always a risk for me to lose my objectivity and demonstrate bias during the interviews. For example, having been both a L2 learner and a L2 teacher in the past, I had my own beliefs, thoughts and opinions concerning teaching and learning English. Sharing these with the participants could easily affect reliability of the data and interfere with the data collection process. To prevent my pre-existing bias from influencing the data in any way, I prepared an interview schedule which had a clear structure with both the main and possible sub-questions and tried to follow the schedule during each interview. All the interview questions were prepared ensuring that none of them was loaded, double-barrelled, or leading questions. The interview schedule was evaluated by my supervisors and revised based on their suggestions. In addition, as mentioned above, the questions were piloted with a number of participants which allowed me to determine the extent to which I sounded neutral. During the interviews, I paid particular attention not to feel over excited or engage with the topics being discussed. I avoided making any comments or using my body language in a particular way (e.g., nodding or frowning) which would be regarded as approving or disapproving what the participants said. Overall, being aware of my own role and position in the current study, I always tried to act as an unbiased interviewer and researcher as much as possible throughout the study.

5.3.3. Data Analysis

The qualitative data was analysed using thematic analysis which is a widely-used qualitative analytic method for organising and describing the data sets (Braun & Clarke, 2006). According to Braun and Clarke (2006), thematic analysis is the foundational method for qualitative analysis as, with the core skills it provides, many other qualitative analyses could be done effectively. Unlike interpretative phenomenological analysis or grounded theory, thematic analysis is not limited to a certain theoretical framework. That is, it can be used with any epistemological approach such as realist and constructionist approaches. This allows researchers to be more flexible with the data analysis as they can both reflect reality by presenting what is in the data, and also go beyond the reality and interpret the data making sense of context (Braun & Clarke, 2006). In the current study, the analysis was informed by critical realism (see Chapter 3 for more information) which emphasises that the reality exists independently from humanity acknowledging that human beings can only acquire the knowledge which is socially constructed and bound to change over time and across cultures.

The analysis was conducted following Braun and Clarke's (2006) six phases of thematic analysis which are as follows: becoming familiar with the data, generating initial codes, searching for themes, reviewing themes, defining and naming themes and producing the final report. Familiarisation with the data was achieved through reading the transcripts line by line a few times and taking notes about the points that could be relevant to the research questions (Phase 1). Following the familiarisation process, the transcripts were systematically coded to determine the recurring semantic similarities across the interviews (Phase 2). The coding was done based on the three issues discussed during the interviews. These were namely the gender stereotyped beliefs about L2 learning, L2 self-efficacy and L2 anxiety. However, the coding process was also open to any further issues discussed by the participants. That is, any additional information gained through the interviews was also included at this stage to provide a more comprehensive and inclusive account of the data. As suggested by Braun and Clarke (2006), the coding was conducted in an iterative way by moving back and forth in the data to achieve a thorough analysis.

Following the coding process, the codes were clustered into the broader themes and a number of sub-themes (Phase 3). The illustrative quotes were also extracted from the transcripts and translated verbatim at this stage, and they were linked to their corresponding themes. The themes were checked against each other to sustain coherence and consistency among them (Phase 4). The overlapping themes were merged into each other and the redundancies were eliminated. To double-check the themes emerged and have a deeper understanding of the data, the transcripts were read for the last time. Once identifying and naming the themes were finalised, the data analysis report which is presented in detail below was produced (Phase 5) (Figure 5.1).

5.4. Results

The analysis focuses on a recurring theme about how gender stereotyping in L2 learning occurred both in and outside language classrooms and the extent to which it was related to success in L2 learning and teaching. The thematic analysis revealed that the overarching theme of the data was supported by two subordinate themes: dispositional factors in L2 learning and individual factors in L2 learning (Figure 5.1).

The subordinate theme of the dispositional factors consisted of four sub-themes which were talent, commitment and effort; sense of responsibility and showing and expressing emotions. As for the individual factors, the data showed that the participants repeatedly

mentioned the link between the learners' past L2 learning experiences, their career aspirations and gender stereotyping of L2 learning. As seen in Figure 5.1., all the sub-themes were informed by a lower order theme of social factors which was concerned with social role expectations and social judgements. This theme was not treated as the main theme as the dispositional and individual factors because it occurred within almost all sub-themes and played a crucial role in explaining the sources of any existing gender stereotyping of L2 learning.

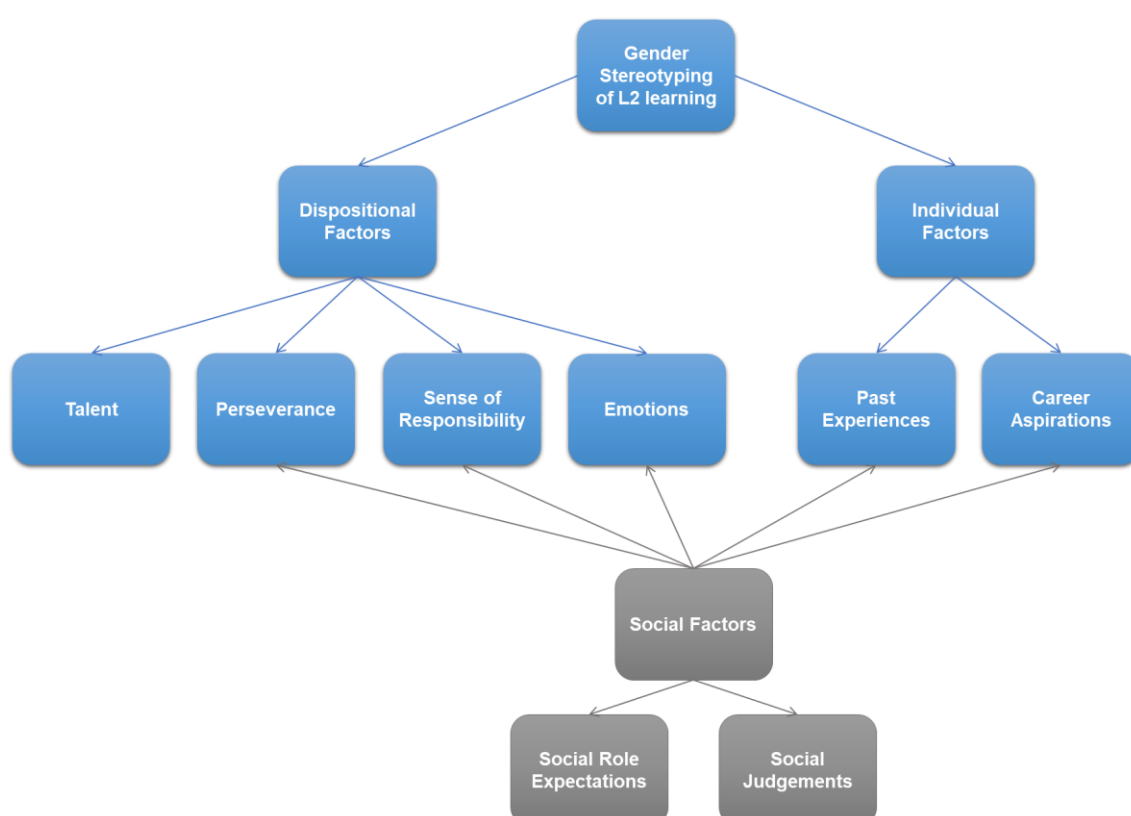


Figure 5.1. *Themes and sub-themes emerged in the thematic data analysis*

5.4.1. Theme 1: Dispositional Factors

5.4.1.1. Innate Talent

Based on the previous literature on gender stereotyping of L2 learning, the participants were asked whether they ever believed females could be more naturally talented in L2 learning than males. The interview data revealed that the participants' views on this matter varied. For example, in contrast to expectations, none of the female learner participants believed that there were any gender differences in L2 learning in terms of talent. They asserted that females and males could be equally talented.

Illustrative Quotations

"I think it is not correct to say that females or males are good at certain things. This changes from person to person. For example, it is believed that females cook better than males, but if you look at the best cooks in Turkey, they are all male. These beliefs are shaped by the society and they are just not correct. People can do anything they want, and they can be successful no matter what gender they are." (Learner, Female).

"I think there are no gender differences in talent. People can achieve whatever they want no matter what their gender is" (Learner, Female).

As the illustrative quotations show, the female learner participants were very careful not to make any strict comments as to whether females and males were more talented in L2 learning. However, unlike the female learner participants, the male learner participants endorsed stronger views that females are better at learning a new language compared to males.

Illustrative Quotations

"Currently, we have 30 EFL learners in the class. I think females are a little bit better than males. For example, when the teacher says something in English, it takes me five seconds to grasp it, but my female classmates understands it immediately" (Learner, Male).

"I think females are more adept at languages than males. It seems like learning English is easier for them, so they might be more successful than males." (Learner, Male).

However, almost half of the male learner participants preferred attributing females' success to some other factors such as interest in L2 learning rather than talent; or treated talent as an acquired asset, rather than an innate ability. That is, they did not prefer crediting females with an innate natural talent which males could not have or achieve even if they wanted to.

Illustrative Quotations

"Females are more talented than males. I don't think I am that talented in English. However, I believe I can improve my talent on my own as well. I need to work five times harder than a female" (Learner, Male).

"I don't think there is a difference between males and females in terms of talent. It is something to do with your interests. You need to have some kind of interest in something so that you might have the opportunity to realize whether you are talented or not. For example, I started attending a drama class two years ago. I didn't like it at all but did not give up easily. I have played more than five roles and realized that I am talented at acting. I can make people laugh. That's why, I think females and males are not different from each other in terms of talent."

They just don't know whether they are talented at anything until they try it. As for L2 learning, females are more interested in this area and they can easily discover their talent" (Learner, Male).

Similar to the female learner participants, the majority of the L2 teachers also indicated that they did not believe females and males were different from each other in terms of talent. According to the teacher participants, talent in L2 learning is an individual difference, rather than an asset which belongs to a particular gender.

Illustrative Quotations

"I don't think we can make such a claim. I have met very talented females and males. I have also met female and male learners which were not talented at all. Based on my experience, I can say that it is not possible to make a certain assumption about this. Success depends on people and their expectations and motivations." (Teacher, Male).

"No, I have had male students as well who were apt to use English effectively. For example, I have recently had a student who will study engineering next year. He is not a social sciences student, so you would not expect him to be talented in L2 learning, but he is unbelievably good at writing in English. I ask him to write one sentence. He writes 15 sentences. He likes expressing himself and this affects the way he uses the language. I don't think success in L2 is well-correlated with gender." (Teacher, female).

However, a closer inspection of the interview data revealed that even if the participants did not have any gender stereotyped beliefs about L2 learners' abilities and talent, they held certain gender stereotyped beliefs about learner behaviours and attitudes towards L2 learning which were mostly informed by the social roles and expectations. These are discussed in detail below.

5.4.1.2. Perseverance

The participants were also asked about their perceptions about any gender differences in L2 self-efficacy. The majority of the participants said that the level of self-efficacy in L2 learning was not related to gender and it changed from person to person. They indicated that they had both female and male learners with higher or lower self-efficacy beliefs. However, across the interviews, there were numerous references made regarding the differences between males and females in terms of perseverance. Several participants noted that males tended to show less perseverance in the face of difficulty. Compared to females, males were more inclined to believe that they would be unsuccessful and give up more easily.

Illustrative Quotation

“Generally, those who are unsuccessful in my classes are males. I think females are better at dealing with failures. At the beginning of the year, even if females are unsuccessful, they keep up the hard work. They do not lose their motivation. They believe that they can achieve. However, males, if they are unsuccessful at the beginning, don’t show any progress. While females put more effort and do their best to succeed, males accept their situation easily. (Teacher, Female).

“Males’ belief in failure is higher in general. Females have strong willpower and when they support their willpower with extra work, they achieve success more easily”. (Teacher, Male).

According to the participants, this was mostly due the social roles attributed to males in Turkey. Some participants mentioned that when males sensed that there was a risk of being unsuccessful in L2 learning, they made the justification that they did not want or need to learn English. As a consequence, they did not put any effort in learning it. In a way, they attempted to protect their self-image as a strong and confident person and did not give their teachers or peers the chance to judge their L2 performance.

Illustrative Quotations

“There is a competition for leadership among males. They would like to be the leader of the whole group, so they don’t want to show their weaknesses. When they don’t succeed enough, they say that “I don’t need to learn English”. Therefore, they believe nobody can say that they are unsuccessful. They want people to think that it is their choice not to continue learning a new language.” (Teacher, Female).

“Due to the social roles attributed to them, males do not want to be seen as a weak person. Therefore, instead of letting people see them failing, they prefer not doing anything at all.” (Teacher, Female).

5.4.1.3. Sense of Responsibility

According to the majority participants, in addition to perseverance, one of the main differences between males and females in L2 learning was their approach to studying and learning. They emphasised that females worked more systematically compared to males. For example, they did their homework in a timely manner and attended the classes regularly. However, males did not seem to be as eager and responsible to perform the additional tasks as females did, which, according to the participants, might result in an achievement gap in the final exam scores between males and females.

Illustrative Quotations

“Females always work more diligently compared to males. They are more motivated and work using appropriate strategies. They are more learning oriented. For example, I have recently had a group of female students. They were all motivated and they were always well-prepared for the lessons. In general, those who are less motivated and skip the classes are males. It is very rare among females. This is evidenced by their exam scores. When I look at the statistics, females are always more successful than males” (Teacher, Male).

“Female students are more responsible. When I look at my students, females are certainly more organised. They take notes and revise at home. They spend extra time on L2 learning. Males don’t.” (Teacher, Female)

Most of the participants indicated that females were more learning and process-oriented whereas males were more outcome-oriented. According to the participants, males would like to achieve success without investing much time in it. They are not as patient as females in L2 learning.

Illustrative Quotations

“To learn a new language, people need to be more patient and work hard as it is a process. I think females are better at working systematically compared to males. However, males are not that patient as they are mostly product oriented.” (Teacher, Male).

“Females are more learning oriented and work more systematically, but males are more pragmatic and work when needed” (Teacher, Female).

Several participants also mentioned that because males preferred learning English in the classroom rather than outside the classroom, they tended to be more active (e.g., answering the questions, volunteering for the activities, raising hands etc.) during the lessons compared to females.

Illustrative Quotations

“Females are better L2 learners, but they are passive and quiet in the classroom. Males are more active.” (Teacher, Female).

“Males prefer doing everything in the classroom. They don’t want to do homework or do any extra work.” (Teacher, Female).

The participants were asked what the potential reasons might be for such differences in attitudes. Several participants mentioned that there were some societal factors affecting the way L2 learners behaved. It was stated that males and females were not treated equally in the society as it was believed that males were superior to females. Being disadvantaged in the society because of their gender, females were used to working diligently to prove their knowledge and abilities.

Illustrative Quotations

"I think females care everything more than males. This is the way we are brought up in the society. As we are in a patriarchal society, males are always flattered. For example, when we are a child, it is said that "It is normal for boys to be naughty" or "You're a girl. You need to be tidy." etc." (Learner, Female).

"This depends on individuals' background. For me, I come from a very relaxed family. My mother and father are like my friends. I had a very relaxed childhood. Therefore, I prefer doing what is easy and tend to care less." (Learner, Male).

5.4.1.4. Emotions

Many participants indicated that females and males also differ from each other in terms of showing and expressing their feelings and emotions. The key point highlighted by the participants was that while females were able to express their emotions such as anxiety freely, males preferred experiencing such emotions themselves. According to the participants, this mainly stemmed from the different social expectations of females and males in Turkey. While males were always expected to be strong and fearless, females were seen as fragile and in need of protection. Several participants also mentioned that when males felt anxious, they had different strategies to protect their self-image including making jokes or being completely silent in the classroom.

Illustrative Quotations

"I think both females and males might feel anxious. It is a matter of individual differences. However, the way they show their emotions might differ. For example, men generally make a joke and try to seem cool and relaxed when they are anxious. (Teacher, Female).

"Males are like indestructible castles in our society. People think that they are strong, so we want to conform to their expectations...Actually, we are anxious as well, but we don't prefer sharing it with people. My friends always say that I am very relaxed, but in fact, I feel anxious from time to time as well. However, I try to hide it somehow. When I am anxious, for example, I make a joke or prefer not to attend the class actively." (Learner, Male).

5.4.2. Theme 2: Individual Factors

5.4.2.1. Past Experiences

As one of the aims of the current study was to investigate the extent to which L2 teachers held gender stereotyped beliefs and explore the link between these beliefs and L2 learning experience, the learner participants were asked their L2 teachers held any stereotypical beliefs or behaviours. From among the learner participants, no one believed

that their language teachers treated males and females in the classroom differently or had different expectations (e.g., females would be more successful) from them. However, they accepted that when they were a high school student, they had some teachers who had such gender stereotyped beliefs.

Illustrative Quotations

“I don’t know how teachers behave in the other universities, but in my university, all language teachers do their best to help us learn English as much as possible. They always try to keep us active and engaged in the classroom. They offer individual support as well. For example, one of our teachers helps me during the classroom activities. She checks my sentences and corrects my mistakes etc. They don’t differentiate between males and females. I think this was in the past. For example, at high school, the language teachers would favour females, but this doesn’t happen here at university” (Learner, Male).

“Our teachers behave equally with each gender. Well, before starting university, I also experienced these stereotypes (like everyone did), such as females are more successful etc. However, we are all mature now and we all have individual differences. I think our teachers are aware of these. Therefore, they don’t expect the achievement level of females and males would be different.” (Learner, Female).

Additionally, some participants mentioned that they avoided producing some sounds in English due to the risk of sounding feminine. In other words, they did not want to damage their ‘masculine’ identity. However, those participants were mostly talking about their experiences as a young learner. None of them referred to their actual L2 learning experience at university level.

Illustrative Quotations

“When I was a student at primary school, I thought L2 learning was associated with femininity. Well, actually, I did not think so. My peers made me believe so. In the language class, I was the only male who was successful. My male friends were always making fun of me whenever they heard me speaking English with a proper accent” (Learner, Male).

“I never thought learning a new language is feminine. However, when I was a student, I might have that impression due to coming from a patriarchal family system. This probably affected the way I pronounced some words and phrases in English, but this was in the past. I do not feel anything like that as a teacher.” (Teacher, Male).

5.4.2.2. Career Aspirations

The majority of the participants disagreed with the notion that ‘L2 learning is a feminine domain’. That is, they did not believe that learning a new language has anything to do with femininity or masculinity. Some of the male participants added that they would not

learn English if they believed it was something associated with femininity. This suggests that although they did not have this stereotyped belief, they would be influenced by gender associations. Additionally, some of the participants highlighted that learning a new language is crucial for everyone no matter what their gender is.

Illustrative Quotations

“It is not a feminine domain, of course. I would not be learning it if it were the case” (Learner, male).

“I don’t think it is a feminine domain and I don’t have any friends who believe so. I think anyone who wants to improve themselves need to learn a new language. This has nothing to do with femininity or masculinity.” (Learner, female).

“it is not feminine. I would not be doing this job if it were so. I think language means communication and everybody needs communication. I enjoy learning languages and I do learn languages in my free time. I can easily say that it is the central part of my life. If someone wanted me to do another job, I would say yes providing that I would use the languages I have learnt” (Teacher, male).

Even though most of the participants disagreed with the role of femininity and masculinity in FL learning, they all stated that they were aware of the fact that females in English majors such English Language Teaching or English Literature outnumbered males in general. Most of the learner participants indicated that they had more female EFL teachers than males throughout their L2 learning experience. However, according to the majority of the participants, female dominance in the field might be attributed to *social orientation*, rather than the nature of L2 learning being feminine or masculine as discussed in the literature. The interviews revealed that there were certain social expectations of females and males in terms of a career path. While females were expected to do jobs that they could easily perform along with their roles as a spouse and mother, males were expected to do jobs that they would earn enough money as a ‘breadwinner’. As the English related occupations were not seen in the latter category, males did not generally prefer pursuing a career in English.

Illustrative Quotations

“I don’t think males don’t choose English as a career path because it is a feminine domain. I think they just don’t consider it as an option because they don’t receive such an input from their environment. In our society, it is common that males are directed at certain directions such as STEM even when they are a child.” (Teacher, Male)

“In Turkey, there are certain jobs which are regarded as more suitable for men than women. For example, families want their sons to be a doctor or an engineer most of the time. I think when a male chooses a career in English, people might think that he was not successful enough to get another job. However, it is not the case for females. They need to find a job that they can easily do.” (Teacher, Male).

The majority of the participants stated being a teacher was regarded as an ideal occupation for a female in Turkey due to its work conditions. Therefore, females chose English majors thinking that they would have the opportunity to be a teacher or do another job that they could easily perform along with their social roles.

Illustrative Quotations

“I think this is mostly because of the work conditions. People think being a teacher is good for females because teachers work half a day. This means that they have time to take care of their children. Unfortunately, we have some certain social roles and are guided accordingly. Families advise their children based on these social roles.” (Teacher, Female).

“My husband is an architect. He works for long hours. He starts working at 8-9 am and works until late. He works even at the weekend. This does not seem to be ideal for a woman especially after she gets married and has children. Unfortunately, some social roles such as taking care of children are given to us. Therefore, women need to choose jobs that are suitable for these roles such as teaching.” (Teacher, Female).

5.5. Discussion

Adopting the qualitative research approach, this chapter explored L2 teachers and L2 learners' perceptions about the common gender stereotypes in L2 learning and the ways in which these stereotypes were constructed. The results suggested that gender stereotyping of L2 learning was mainly concerned with two broad themes: dispositional factors and individual factors. As discussed in Chapter 1, the differences between Turkish male and female EFL learners should be considered in the context of Turkish societal values. Considering females and males have quite defined gender role expectations in Turkey, it is highly likely that these gender role expectations are reflected in L2 learning process. These are discussed in detail below.

5.5.1. Dispositional Factors

In line with the earlier findings (e.g., Carr & Pauwels, 2006; Plante et al., 2009), the results of this study showed that there were some certain gender stereotypes pertaining to L2 learning among the participants. However, these stereotypes do not seem to be about language learners' abilities as such. They are mostly about the attitudes and

behaviours towards L2 learning, and this might be hindering the learning process in some respects. As for talent in languages, it was a general consensus from both learners and teachers that both males and females might be talented in L2 learning. This finding confirms van der Vleuten's (2016) assertion that the existing gender stereotypes in L2 learning are not due to the gender differences in ability or talent. They mostly stem from the attitudes regarding the appropriate roles, behaviours and responsibilities of males and females in a particular society.

In relation to the attitudes towards L2 learning, this study found that there are certain differences between males and females in terms of displaying perseverance in the face of difficulty. The participants argued that learning a new language is a process and it takes some perseverance and hard work. Compared to females, males were revealed to be less persevering in L2 learning. Interestingly, this does not align well with the normative roles associated with males in Turkey. According to Boratay, Fisek and Ziya (2014), being a 'man' in Turkey is associated with hard work, toughness, endurance, determinedness, success and power. As the breadwinner of the family, males are more powerful than females and have the final say in important decisions for the family members. However, when it comes to L2 learning, males seemingly do not show as much resilience as females when faced with difficulties. As some of the participants highlighted, a possible explanation for the aforementioned differences between males and females might be that males aim at protecting their self-image. In Turkish society, males are mostly associated with strength and confidence. Therefore, in the context of L2 learning, instead of taking the risk of failing and damaging their 'male' figure, males prefer avoiding the tasks at which they might not be as successful as they desire to be. Unlike males, females are more likely to evaluate their weaknesses and address them more effectively to achieve better results in L2 learning. Therefore, in the face of difficulty, it is highly likely that females persevere more and achieve higher levels of L2 proficiency compared to males.

Such findings have important implications for developing a resilient sense of efficacy in L2 learning. As discussed in Chapter 2, experiencing some difficulties and setbacks enable people to acquire a resilient sense of efficacy (Bandura, 1997). When people achieve successes easily, they might assume that it will be case in any future tasks, and therefore, can lose their courage when faced with failure. In contrast, they learn how to deal with failures after gaining some experience in overcoming obstacles. Given that female EFL learners tend to persevere more compared to males and do not give up

easily even when faced with difficulties, it is highly likely that they develop higher self-efficacy in L2 learning compared to males.

Another important finding concerning attitudes was that females were revealed to be more hardworking than males in L2 learning. This finding mirrors those of the previous studies that have investigated the gender differences in L2 learning (e.g., Carr & Pauwels, 2006; Lu & Luk, 2014). The participants reported that females were more likely to be engaged in extra work such as doing homework and taking notes during the classes. They were also more organised and had a more systematic approach to studying a new language. In contrast, males tended to be social and active during the classroom activities and were found to be less patient and reluctant to take further responsibilities outside the classroom. Since L2 learning requires ongoing and sustained involvement rather than just participating in the lesson, such attitudes might have a negative impact in L2 male learners' language learning experience. The findings concerned with the differences between males and females in terms of the attitudes may be explained by the fact that females and males are expected to perform different social roles in Turkey. According to Uğurlu et al. (2018), while females are expected to be obedient, family-oriented and responsible, males are generally associated with dominance, independence, and aggression. The results suggest that most of the learners internalised these gender role expectations and behaved accordingly while learning a new language. Such differences in behaviours have an important implication for L2 teaching and learning. Even though L2 teachers provide learners with many learning opportunities in language classrooms, they are not always adequate to be a proficient in the target language. Considering that L2 learning is a long and complex process, L2 learners need to invest their time in this process over and beyond the opportunities given in the classroom. Undertaking extra work while learning languages provides learners with more opportunities to identify their knowledge gaps and bridge these gaps timely. Based on the findings of this study, it can be assumed that since female learners tend to invest their time in this process more, they might be in a more advantageous position compared to male learners. Similarly, since female learners had more belief that they could master the subject whilst males held entity views on intelligence, it is less likely for males to improve compared to females having the same language proficiency.

The last dispositional factor was the difference between males and females in terms of showing and expressing their emotions and feelings. For example, it was generally agreed that females were more prone to demonstrate their anxiety. In contrast, males

tended to hide their anxiety and handle it adopting different strategies such as making jokes. Similar to the issue of perseverance, it seems possible that these results are due to males' intent to protect their self-image. In Turkey, it is expected from males to be dominant, confident and strong. As such, even if they feel anxious or lose their confidence in L2 learning, they cannot express or demonstrate these feelings as freely as females do because they are afraid of looking weak and vulnerable. Such an attitude might be hindering the learning process for males in L2 learning in two ways. First, males' being protective of their self-image might result in their deprivation of essential teacher support. Across the interviews, several learner participants mentioned that their teachers' emotional supports played a crucial role in managing their anxiety. This was also reiterated by the teacher participants who indicated that they adopted different approaches and strategies to help anxious learners. However, when learners do not show their feelings or ask for help explicitly, it becomes more challenging for L2 teachers to recognise anxiety in learners and offer help accordingly. Secondly, since males do not want to expose their weaknesses, it is less likely that they ask questions when they do not understand anything in class. Being afraid of appearing 'less knowledgeable' in front of their teacher and peers, they might not benefit from the opportunities available to full extent. As such, there might be more gaps in their L2 knowledge compared to females.

5.5.2. Individual Factors

In addition to the dispositional factors, the participants also mentioned some individual factors which might results in learners' and teachers' gender stereotypic approach to L2 teaching and learning. These were grouped under two sub-themes: learners' past L2 learning experiences and choosing L2 as a career path. As part of the research aims, the participants were asked whether their L2 teachers played a role in sustaining or legitimising the common gender stereotyped beliefs. The findings suggested that L2 teachers at university level, even if they sometimes agreed with some of the gender stereotyped beliefs themselves, did not reflect these beliefs to the learners. This was also confirmed by the L2 learner participants who stated that they did not see any differences in their L2 teachers' behaviours towards females and males. Therefore, one possible conclusion is that L2 teachers teaching at university level do not appear to explicitly sustain or legitimise the existing gender stereotyped beliefs pertaining to L2 learning. As such, these beliefs, if any, might be created and shaped prior to university. That is, L2 teachers do not appear to make the issue of gender stereotypes salient and do not account for any observed differences in any of the points investigated in the current study, but there are wider factors such as societal values which seem to be more influential. This was partly evidenced in the findings of this study. For example, some of

the learner participants shared their experiences either with their previous L2 teachers or their family members who adopted a gender stereotypic approach towards L2 learning. This finding, while preliminary, suggests that both primary and secondary schools and the patriarchal family structure might play a crucial role in learners' gender stereotyped beliefs.

The results also revealed that even though the participants had more female L2 teachers than males, they still did not believe that L2 learning is a feminine domain. Most of the participants thought that the main reason for the female dominance in the field was not because they were more talented or had superior abilities in L2 learning than males. It was mostly associated with the gender roles attributed to females in the society. As mentioned above, females in Turkey are generally family-oriented and they are expected to choose jobs that they can perform these roles appropriately. Believing that being a teacher is suitable for females, most females pursue a career in this field. It can thus be suggested that it is the teaching profession which is feminized by L2 learners, not the L2 learning per se.

5.5.3. Limitations and Suggestions for Future Research

The above results must be interpreted with caution because the generalisability of them is subject to two main limitations. First, this study was conducted among adult L2 learners learning English on a compulsory basis. Therefore, these findings cannot be extrapolated to L2 learners in different contexts such as those learning a new language on a voluntary basis. Therefore, it can be said that there is abundant room for further progress in determining the extent to which gender stereotyping of L2 learning occurs in language classrooms and whether it is linked to L2 learners' achievement. Also, the results of the current study suggested that gender stereotypes might be prevalent in primary and secondary schools in Turkey. As it was beyond the scope and intent of the current research, this area was not explored in detail in the current research. Therefore, future research needs to further examine these areas to address this limitation. Particularly, this may help establish a more refined idea of at what point gender socialisation and gender role expectations become prominent in a learner's behaviours and learning potential.

Secondly, although adopting a qualitative research approach has a number of advantages such as offering rich and in-depth data, it also brings about some disadvantages. As Creswell (2014) points out, interviews provide subjective rather than objective data. It is not always possible to ascertain the extent to which participants share

their exact thoughts and feelings with researchers. Therefore, it is suggested that interview data is triangulated adopting some other data collection methods. A further study with classroom observations, for example, could provide a more complete picture of the phenomenon in question. Also, due to the nature of the qualitative research methodology, it can only be assumed that certain variables might be related to each other. That is, no causation can be inferred from the data available. Researchers seeking for a causal explanation should adopt a different methodology and approach the issue of gender stereotyping of L2 learning accordingly.

It is also important to note that the interview questions in the current study were mostly concerned with the differences between males and females in terms of FL learning. It is possible that these questions could have primed the report of gender differences rather than similarities. Therefore, further research should be done to investigate the similarities as well to provide a more balanced understanding of the extent to which FL learning is believed to be a female domain.

5.5.4. Original Contribution to Knowledge

This study makes three contributions to the current literature. First, it is among few studies investigating adult L2 learners' perceptions about gender stereotyping of L2 learning in a non-English speaking country. As the research on gender stereotyping in L2 learning has been mostly limited to primary or secondary school students in English-speaking countries, the empirical findings in this study provides additional evidence with respect to the phenomenon of gender stereotyping of L2 learning. Secondly, as previous studies concerning gender stereotyping of academic subjects have mainly focused on females and their academic achievement in certain so-called masculine academic fields such as STEM, males and their achievement in so-called feminine academic fields are understudied, especially in the field of language studies. The current study, therefore, proposes that the issue of gender stereotyping of academic subjects does not only concern females, but also males and sheds light on this issue in context of L2 learning. Lastly, the study has gone some way towards enhancing our understanding of the sources of any existing gender stereotypes pertaining to L2 learning. It offers some important insights into the extent to which L2 learning environment fosters the salience of gender-stereotyped beliefs and explores whether L2 teachers as an agent of socialisation in the learning environment play a role in sustaining and legitimising these beliefs.

5.6. Chapter Summary

Chapter 5 has presented the second empirical study of this thesis which qualitatively investigated L2 learners' and teachers' thoughts and beliefs about three main concepts: gender stereotyped beliefs in L2 learning, L2 self-efficacy and L2 anxiety. The results indicated that although the participants did not explicitly believe that L2 learning was a feminine domain, they accepted that there were certain dispositional and individual differences between males and females in L2 learning. The findings also demonstrated that L2 teachers did not play a role in sustaining and legitimising any existing gender stereotypes that L2 learners held. The results of this study are highlighted further in Chapter 7 which integrates the findings of all the three studies and discuss them with references to the research literature and research questions framing this thesis.

CHAPTER 6: STUDY 3

6.0. Chapter Outline

The purpose of the current chapter is to provide and discuss the findings gained from Study 3, which adopted an experimental design. Differently from the previous two studies, Study 3 experimentally investigates the impact of stereotype threat pertaining to learning another language upon L2 learners' language performance via the mediating roles of L2 anxiety and L2 self-efficacy. The chapter is divided into three main sections. The first section focuses on the gaps in the stereotype threat literature and establishes the significance of the current study in addressing these limitations. The second section describes the experimental design and procedure in detail and shares the results gained from the experiments. Finally, the discussion and conclusion present an overview of the findings. These are followed by the limitations of the study and the implications for further research.

6.1. Introduction

Stereotype threat (ST) which refers to "being at risk of confirming, as self-characteristic, a negative stereotype about one's group" (Steele & Aronson, 1995, p.797) has gained a great deal of interest and momentum in the research literature since its first appearance in social psychological research (Inzlicht & Schmader, 2012; Picho et al., 2013). According to Google Scholar's citation report accessed in 2018, the seminal article on this phenomenon by Steele and Aronson (1995) has been cited approximately 8,500 times. Inzlicht and Schmader (2012) state that the popularity of ST is mostly because it offers an explanation for group differences in academic performance. ST avoids explaining such differences based on nature versus nurture debate. It suggests that immediate situations themselves can lead to apparent group differences in academic performance (Inzlicht & Schmader, 2012; Steele, 1997). This situational approach is significant in that the academic performance of stigmatised groups can be increased by using some simple interventions which are designed to remove these threats (Inzlicht & Schmader, 2012). As Inzlicht and Schmader (2012) advise, it is not possible to change people's biology or the way they are brought up. Instead, ST suggests that it is possible to change the situations people find themselves in and consequently increase their performance.

ST has inspired several researchers from different disciplines and Steele and Aronson's (1995) original study has been replicated many times with varying demographics in a variety of performance domains including; older adults and memory (Barber & Mather,

2013; Levy, 1996), women and science, technology, engineering, mathematics (STEM) (Picho et al., 2013), women and leadership (Hoyt & Murphy, 2016), white males and athleticism (Stone et al., 1999); and women and gaming (Kaye & Pennington, 2016). Although there are several stereotypes pertaining to foreign language learning such as L2 learning is a female domain (see Carr & Pauwels, 2006; Schmenk, 2004), little research has been carried out to explore these within the ST context (Paladino et al., 2009). The current study therefore aims to make a theoretical contribution to knowledge by answering the following questions posed by Schmenk (2004), which have not been fully addressed to date:

“Does the phenomenon called *stereotype threat* in social psychology (Kray, Thompson, & Galinsky, 2001; Steele, 1997), the tendency to confirm gender stereotypes when they are explicitly activated, occur in language learning contexts as well? What role do stereotyping and so-called self-fulfilling prophecies play in sustaining the belief in language learning as a feminine domain?...” (p.521).

Also, the initial studies in ST conceptualised stereotype threat as a single construct which, according to Shapiro and Neuberg (2007), resulted in some inconsistency in the ST literature and therefore, made the generalisability of much published research on ST problematic. For example, Aronson et al. (1999) question whether ST is self-threatening because individuals are afraid of being a bad ambassador of their group to the society or because they are simply worried about their own reputation and are afraid of being incompetent in the eyes of other people (Aronson et al., 1999). Challenging the idea of ST being a single construct, Shapiro and Neuberg (2007) have suggested the multi-threat framework which was discussed in Chapter 2 at length. This 2x3 framework posits that there are six types of stereotype threat which are categorised according to the source (the self, the in-group, or the out-group) and the target of the threat (the self or the in-group).

As Paladino et al. (2009) state some people might be more susceptible to one type of threat than others and they might show lower performance depending on which type of threat is made salient. In the present study, the focus is on the target of the ST (i.e., who does the action reflects upon: the self or one's group) because, as Shapiro et al. (2013) point out, studies in the ST literature has mainly examined either self-as target or group-as-target stereotype threat conditions although they do not explicitly differentiate between these two types. In these studies, ST is conceptualised either as a concern about representing the group or a concern about how a performance might be a representation of one's stereotypical abilities. As such, the distinction between self-as-target and group-as-target stereotype threats is naturally evident from the existing

research. To contribute to the growing literature, one of the aims of this study is to determine whether self-as-target (i.e., I am afraid that my behaviour will confirm, *in my own mind*, that the negative stereotypes which are held of my group are *true of me*) or group-as-target (i.e., I am afraid that my behaviour will confirm, *in my own mind*, that the negative stereotypes which are held of my group are *true of my group*) is more detrimental to Turkish males' L2 performance.

The multi-threat framework also suggests that in addition to the distinct stereotype threats, there are also various mechanisms which play a crucial role in determining whether ST has a greater or lesser effect on individuals' performance. These mechanisms have been investigated by several researchers. In their recent systematic literature review, for example, Pennington et al. (2016) identified a variety of different affective, cognitive and motivational processes that might explain the effects of stereotype threat on different kinds of performance outcomes. Among the potential affective mediators are anxiety and self-efficacy which are the main interests in the current research. According to Pennington et al. (2016), the findings concerning whether ST affects performance through anxiety and self-efficacy are not consistent across the studies. For example, in Steele and Aronson' (1995) original study, self-reported anxiety did not mediate the ST effect on African American's intellectual performance whereas Chung et al. (2010) revealed that both self-reported state anxiety and specific self-efficacy were the mediators of the effects of stereotype threat on African American's promotional exam performance. Unlike Chung et al.'s (2010) study, in some other studies, self-efficacy did not mediate the effects of self-as-target threat on African American's cognitive ability (Mayer & Hanges, 2003) or group-as-target threat on women's mathematical performance (Spencer, Steele, & Quinn, 1999). These findings suggest that not all individuals are affected by ST in the same way.

The aforementioned mixed findings might be attributed to a number of other variables which are believed to moderate the effects of stereotype threat on the performance. For example, previous research has examined the role of stereotype endorsement in targeted individuals' performance (Pennington et al., 2016). Schmader et al. (2004) found that females who endorsed the belief that maths is a male domain had negative self-perceptions about their maths competence and expressed less interest in pursuing a career in the field. Therefore, it can be suggested that those who endorse the common negative stereotypes attributed to the group they belong to might be more affected by ST than the others. To test this hypothesis in the current study, whether the level of

stereotype endorsement pertaining to female superiority in language learning moderates the effects of ST on L2 male learners' L2 anxiety and L2 self-efficacy is also explored.

It is also theorised that ST effects tend to emerge on tasks of high difficulty and demand (Pennington et al., 2016). For example, Keller (2007) found that ST had different effects on low and high domain identifiers' (i.e., maths is important to me) performance depending on the task difficulty. When the task was difficult, low identifiers showed higher performance whereas high identifiers had lower performance under ST. However, such results did not emerge when the task was easy. While acknowledging that each language skill (i.e., listening, speaking, reading and writing) is an integral part of L2 learning and they all have their own challenges, the current study focuses on an L2 listening task for three main reasons. First, it is the least studied skill which has not gained attention until recently (Goh, 2016). It is evident that more research is needed to develop a better understanding of this language skill. Secondly, L2 listening has a crucial role in L2 learning. Not only does it have a role in communication, but it is also an essential source of language input and provides a number of opportunities for L2 development (Goh, 2016). Therefore, the task value of L2 listening is very high. Lastly, compared to other skills, L2 listening is more challenging for some L2 learners (Brown, 2016; Goh, 2016). Studying L2 listeners' challenges from a cognitive perspective, Goh (2000) found that L2 listeners experience comprehension problems broadly in three categories: perception (e.g., failing to recognise the words learnt before or experiencing concentration issues), parsing (e.g., forgetting what is just heard), and utilisation (e.g., failing to understand the intended message in the listening text). There are also some other external factors influencing L2 listeners' achievement such as speech rate, unfamiliar accents or noise in the environment (Brown, 2016). Considering these challenges, a L2 listening task is potentially of high difficulty and demand which might easily heighten any ST effects.

In light of the previous studies and their findings, Study 3 aims to fill in the aforementioned gaps and make a significant contribution to the limited ST literature in the field of foreign language learning with a special focus on L2 listening. To achieve this, a number of research aims and questions were created. These are presented in the section below.

6.2. Research Aims and Questions of Study 3

Study 3 experimentally investigated the effect of stereotype threat with respect to foreign language learning on male learners' L2 listening performance through the mediating roles of L2 listening anxiety and L2 reception self-efficacy. The research questions of the current study were as follows:

1. To what extent does gender stereotype threat pertaining to foreign language learning affect male learners' L2 listening performance via L2 listening anxiety and L2 reception self-efficacy?
 - 1.1. Does self-as-target and group-as-target stereotype threat influence differently males' L2 listening performance?
 - 1.2. To what extent does L2 listening anxiety mediate the effects of self-as-target and group-as-target stereotype threat?
 - 1.3. To what extent does L2 reception self-efficacy mediate the effects of self-as-target and group-as-target stereotype threat?

6.3. Method

6.3.1. Participants

A total of 103 Turkish male learners studying EFL at university level were recruited ($M_{age} = 19.23$, $SD = 1.17$). Participants were assigned randomly to a self-as-target ($n = 42$) or group-as-target stereotype threat condition ($n = 33$) or to a control condition ($n = 26$). In addition to these, there were also two participants whose data was excluded because they did not complete the pre and/or post questionnaires required as part of the study.

6.3.2. Stereotype Threat Manipulations

Adopting Shapiro and Neuberg's (2007) multi-threat framework, participants in the treatment conditions were primed with either a self-as-target or group-as-target stereotype threat. The priming was carried out twice: before the questionnaires and before the performance measure (i.e., English listening test), and it was in Turkish. The stereotype manipulations implemented in each condition are described in detail below.

6.3.2.1. Self-as-target stereotype threat

Male participants in the self-as-target condition were primed with a negative gender stereotype concerning their ability in L2 learning compared to their female peers. The following instructions were designed to make the gender stereotype that L2 learning is a female domain salient and emphasise that the test was diagnostic of the L2 learners' individual ability as a male, not that of their group's ability as a whole. The instructions they received were as follows:

"There is a negative stereotype that males have less ability in foreign language learning than females. Today, I would like you to take an English listening test. The score you will get from this test will help me establish your personal ability in L2 learning as a male."

6.3.2.2. Group-as-target stereotype threat

Participants in the group-as-target condition were primed with the same gender stereotype. However, unlike the participants in the self-as-target stereotype threat condition, they were informed that the test was diagnostic of their gender group, not that of their individual performance. The below instructions aimed to heighten the salience of gender stereotype and emphasise that it targeted the males' ability as a gender group:

"There is a negative stereotype that males have less ability in foreign language learning than females. Today, I would like you to take an English listening test. The score you will get from this test will help me establish males' ability in L2 learning in general."

6.3.2.3. Control condition

Participants in the control group did not receive any information about the negative gender stereotype shared with the participants in the treatment conditions. They were informed that the test was not diagnostic of their ability. They were primed with the instructions below:

"In today's session, I want to assess the reliability of this listening test which I have prepared for foreign language learners. I would like to see how many correct and wrong answers each question will get. Therefore, this English listening test is not diagnostic of your language learning ability. Please listen to the audio recordings carefully and answer the questions accordingly."

6.3.3. Data Collection Instruments

6.3.3.1. Background Questionnaire

A background questionnaire was designed to obtain information about the participants' demographic information (e.g., age, home region) and their L2 learning experiences (e.g., duration of English study, languages spoken other than English). The participants completed the background questionnaire before the stereotype manipulation.

6.3.3.2. L2 Listening Anxiety Sub-Scale of the MLCAS

To assess participants' L2 listening anxiety, the related sub-scale of the MLCAS was used (see Chapter 4 for the detailed information). In this questionnaire, there were six items (i.e., two items for each affective, cognitive and physiological components). The participants were asked to rate the items on a 5-point scale (1 = strongly disagree, 5 = strongly agree). The reported Cronbach's alpha for this specific sub-scale was .89. The participants completed L2 Listening Anxiety Sub-Scale of the MLCAS after the stereotype manipulation, but before the English listening test.

6.3.3.3. L2 Reception Self-efficacy Sub-scale of the QLLS

Participants' listening self-efficacy was measured using the reception part of the QLLS (see Chapter 4 for more information). There was a total of five items in this questionnaire which were rated on a scale of 1 to 5 (1 = strongly disagree, 5 = strongly agree). Reliability analysis of this sub-scale revealed a Cronbach's alpha value of .85. The participants completed L2 Reception Self-efficacy Sub-scale of the QLLS after the stereotype manipulation, but before the English listening test.

6.3.3.4. English Listening Test

Participants' English listening proficiency was assessed using one of the Preliminary English Tests (PET) designed by Cambridge Assessment English (UCLES, 2015) which is part of Cambridge Assessment Group, a non-teaching department of the University of Cambridge. Cambridge Assessment English is well-known for its tests and exams which are recognised around the world by many employers, universities and organisations as proof of ability to use English. Therefore, the reliability and validity of the tests prepared by this well-established institution are high.

The PET was addressed to L2 learners at the pre-intermediate (B1) level of the CEFR (see Section 4.3.3.3 for the descriptions of the levels) and aimed to show whether language learners mastered the basics of English and had practical language skills for everyday use. To avoid the ceiling effect on performance (i.e., that the test is either too easy or difficult resulting in similar scores in performance), the level of the test was just above the participants' actual English level. At the time of the data collection, the participants had been learning for about six months suggesting that most of them would have already completed the A1 and A2 levels and had just started studying at B1 level. Prior to the experiments, the level of the participants was double-checked with the language teachers of the chosen classrooms and a particular attention was given not to select any participants who were at A1/A2 or B2/C1 levels.

A PET exam is originally designed to assess all four language skills (i.e., listening, reading, writing, speaking) and lasts about 2 hours 20 minutes. As the main focus of the current study was L2 listening, only the relevant part of the exam was taken. The listening part of the PET consisted of 4 sections and it took 12-13 minutes to complete them. The details of the sections are as follows:

- In part 1, the participants listened to seven short recordings. For each recording, they were asked to choose the best of three pictures.
- In part 2, the participants listened to a longer recording and answered 6 multiple choice questions accordingly.

- In part 3, there was a longer monologue. While the participants listened to the recording, they needed to complete the missing information in the given text. There were six gaps to complete.
- In part 4, the participants listened to a longer recording and answered 6 True/False questions accordingly (see Appendix 6.1. for a copy of English listening test).

6.3.3.5. Gender Stereotype Endorsement

As one of the aims of the current study was to investigate the extent to which the level of stereotype endorsement moderated the impact of L2 listening anxiety and L2 reception self-efficacy on the participants' performance, two questionnaires were adopted to determine the extent to which the participants agree with the common negative stereotypes concerning males' L2 ability.

Part 1: Participants' Gender Stereotyped Perceptions of Subjects

In this questionnaire, the participants were asked to rate a series of 16 academic subjects which included English as well, on a 5-point scale (1 = females always better, 5 = males always better). The midpoint of 3 was neutral (no difference). The questionnaire aimed to provide a wider picture of the participants' existing gender stereotyped beliefs about the common academic subjects. This part was completed by the participants before the stereotype manipulation.

Part 2: Questionnaire of Language Learners' Gender Stereotyped Beliefs

This questionnaire specifically focused on L2 learning. It comprised of two sections. The first section consisted of five statements assessing EFL learners' gender stereotyped beliefs about foreign language learning. The second section was made up of the same items, but they aimed to measure EFL learners' perceptions about their L2 teachers' gender stereotyped beliefs. The items were rated using a 5-point scale (1 = females always better, 5 = males always better). As mentioned earlier, the midpoint of 3 was neutral. The Cronbach's alphas for these were .75 and .79 respectively (see Chapter 4 for more information). This part was completed by the participants after the English listening test.

6.4. Procedure

Participants were recruited from among the EFL learners at B1 level (independent user) using the convenience sampling method. The study was conducted in the classrooms where the participants' EFL education normally took place. The permission was granted

by the directors of the Foreign Languages Schools and the teachers responsible for the selected classes. The learners were also asked to give their individual consent on the first page of the questionnaires. Each experiment lasted 40-45 minutes in total. After giving information about the nature of the study and getting the participants' consent, the participants were asked to fill in the background and Participants' Gender Stereotyped Perceptions of Subjects questionnaires. On completion of these questionnaires, the participants were randomly assigned to one of the three conditions (self-as-target, group-as-target and control). Based on their condition, they received the relevant instructions described in Section 6.3.3 which was followed by the questionnaires of L2 listening anxiety and L2 reception self-efficacy. Once the questionnaires were completed, the participants were primed with the same instructions again and asked to complete the English listening task. The participants were informed that they would hear the listening text only once, so they might need to be ready to take notes or use their own strategies to answer the questions. Following the task, the participants were asked to complete the Questionnaire of Language Learners' Gender Stereotyped Beliefs which aimed to determine whether they agreed with the negative stereotypes pertaining to the males and L2 learning. At the end of each experiment, the participants were thanked for their participation and were provided with a debrief which emphasised that the negative stereotypes mentioned at the beginning of the experiment were not a true reflection of their ability.

6.5. Results

6.5.1. Descriptive Results

The demographic information about the participants was collated using the details provided in the background questionnaire. All the information was analysed using SPSS v.24. Table 6.1 below presents the analysis of the demographic information and the participants' background in EFL learning. Most of the participants were from Marmara Region (65%). The rest of the participants were from Black Sea Region (2.9%); Aegean Region (8.7%); Mediterranean Region (7.8%); Central Anatolia Region (5.8%); South-eastern (3.9%); and Eastern Anatolia Region (1.9%).

The analysis revealed that only 12.6% of the participants travelled abroad before. The majority of the participants (87.4%) had limited exposure to the target language outside the language classrooms. At the point of data collection, 91.3% of the participants had been studying EFL for 6-12 months at university. 62.1% of the participants declared that they had studied English before. However, 84.5% of these participants believed that their

previous English education was not enough or good quality. Also, almost 80% of the participants indicated that they did not learn any other foreign language except English.

Table 6.1. Frequency analysis of demographics and foreign language background of the participants - Experimental Study

Demographic		N	Percentage
Gender	Male	103	100%
	Female	0	0%
Age	18-25 Years	94	91.3%
	Missing	9	8.7%
Home Region	Black Sea Region	3	2.9%
	Marmara Region	67	65%
	Aegean Region	9	8.7%
	Mediterranean Region	8	3.9%
	Central Anatolia Region	6	5.8%
	South-eastern Anatolia Region	4	3.9%
	Eastern Anatolia Region	2	1.9 %
	Missing	4	3.9%
Duration of English study at university?	6-12 months	94	91.3%
	> 12 months	7	6.8%
	Missing	2	1.9%
Foreign Language other than English?	No	82	79.6%
	Yes	21	20.38%
Been abroad?	Yes	13	12.6%
	No	90	87.4%
English Study before university?	< 12 months	3	2.9%
	1-5 years	41	39.8%
	5-10 years	19	18.4%
	> 10 years	31	30.1%
	Missing	9	8.7%
Quality of English Study before university	Definitely Enough	4	3.9%
	Enough	9	8.7%
	Not Enough	31	30.1%
	Definitely not Enough	56	54.4%
	Missing	3	2.9%

6.5.2. Gender Stereotype Endorsement

A series of one-sample *t*-tests were conducted to determine the extent to which male L2 learners endorsed the gender-related stereotypes attributed to the given academic subjects. The participants were asked to rate these subjects on a 5-point scale (1 = Females are always better, 5 = Males are always better). The test value of 3 (i.e., Both males and females are equally better) was defined as the neutral value. Table 6.2. presents the results gained from the one-sample *t*-test analyses.

Table 6.2. Descriptive Statistics of Participants' Gender Stereotyped Perceptions of Academic Subjects

	N	t	df	M	M difference	p
1. Fine arts.	101	-.84	100	2.94	-.059	.399
2. Educational Sciences.	100	1.86	99	3.14	.140	.066
3. Sports Sciences.	101	12.48	100	3.86	.861	.000
4. Aviation and Space Sciences.	101	12.70	100	3.97	.970	.000
5. Architecture and Design.	101	-4.73	100	2.64	-.356	.000
6. Economics and Administrative Sciences.	101	5.06	100	3.44	.436	.000
7. Medical Sciences.	101	-.51	100	2.96	-.040	.608
8. Dentistry.	101	-1.29	100	2.90	-.099	.198
9. Health Sciences.	101	-13.15	100	2.04	-.960	.000
10. Life Sciences.	100	4.05	99	3.32	.320	.000
11. Humanities.	101	-.123	100	2.99	-.010	.902
12. English.	101	-2.70	100	2.80	-.198	.008
13. Engineering.	101	13.85	100	3.89	.891	.000
14. Communication.	101	-4.23	100	2.67	-.327	.000
15. Tourism.	101	.266	100	3.02	-.059	.791
16. Agriculture.	101	8.24	100	3.60	.140	.000

As shown in Table 6.2, the results concerning seven academic subjects (i.e., fine arts, educational sciences, medical sciences, dentistry, humanities and tourism) were non-significant (all *ps* > .05). Consistent with the literature and the expectations, the male L2 learners endorsed the gender stereotype that females are better at foreign language learning than males $t(100) = -2.70$, $p < .01$. Additionally, the participants believed that females are better than males at three other academic subjects, namely architecture and design: $t(100) = -4.73$, $p < .001$, health sciences: $t(100) = -13.15$, $p < .001$ and

communication: $t(100) = -4.23, p < .001$. As for the academic subjects which were believed to be a male domain, the participants endorsed that males are better than females in sports sciences: $t(91) = 4.86, p < .001$, aviation and space sciences: $t(91) = 4.86, p < .001$, economics: $t(91) = 4.86, p < .001$, life sciences: $t(91) = 4.86, p < .001$, engineering: $t(91) = 4.86, p < .001$ and agriculture: $t(91) = 4.86, p < .001$.

Table 6.3. Univariate Test Results – Gender Stereotype Endorsement

	<i>F</i>	<i>df</i>	<i>p</i>	η_p^2
1. Fine arts.	1.56	2, 97	.215	.031
2. Educational Sciences.	.358	2, 97	.700	.007
3. Sports Sciences.	4.15	2, 97	.019	.079
4. Aviation and Space Sciences.	2.75	2, 97	.068	.054
5. Architecture and Design.	1.09	2, 97	.339	.022
6. Economics and Administrative Sciences.	2.25	2, 97	.110	.044
7. Medical Sciences.	.960	2, 97	.387	.019
8. Dentistry.	.500	2, 97	.608	.010
9. Health Sciences.	1.37	2, 97	.258	.028
10. Life Sciences.	1.68	2, 97	.190	.034
11. Humanities.	.262	2, 97	.770	.005
12. English.	.094	2, 97	.910	.002
13. Engineering.	2.02	2, 97	.137	.040
14. Communication.	1.07	2, 97	.347	.022
15. Tourism.	1.74	2, 97	.180	.035
16. Agriculture.	1.27	2, 97	.285	.026

Also, a series of MANOVAs were conducted to determine the extent to which male L2 learners in the three conditions differed from each other in terms of the endorsement level. The overall MANOVA analysis revealed that there were not significant differences between the conditions, $F(32,166) = 1.26, p = .175, \eta_p^2 = .196$. Although it is suggested that a non-significant MANOVA result should not be followed up with univariate tests, the below analyses were conducted and added in this section for exploratory purposes only. Table 6.3. presents the results gained from the MANOVAs. As shown in Table 6.2, there were not any significant differences across the three conditions in any academic subjects except for the sports sciences. The perceptions of the self-as-target group ($M = 4.10, SD = .718$) were significantly different from the group-as-target group ($M = 3.69, SD = .718$).

The control group ($M = 3.73$, $SD = .132$), however, did not significantly differ from the self-as-target condition or the group-as-target group ($M = 3.68$, $SD = .114$). The results gained from this pre-test measure suggested that the participants' pre-existing beliefs were controlled. That is, any differences in the results were due to the stereotype manipulation rather than the participants' existing beliefs. As the only difference was sports sciences, it is unlikely that this belief had any bearing on English-related beliefs and performance.

6.5.3. Stereotype Manipulation Check – L2 Learning

To evaluate whether there were any differences in the level of stereotype endorsement across the three condition groups, a multivariate Analysis of Variance (MANOVA) was conducted in SPSS. Unexpectedly, the analysis revealed that the stereotype manipulation used in the current study was not successful, $F(4,190) = .782$, $p = .538$, $\eta_p^2 = .016$. As highlighted above, such a non-significant MANOVA result suggests that no follow-up analyses should be conducted. Therefore, it needs to be acknowledged here that the below analyses were conducted and added in this section for exploratory purposes only.

The univariate test results are also presented in Table 6.2 below. The results suggest that the group presented with the self-as-target threat did not hold significantly different gender stereotypical beliefs concerning L2 learning ($M = 14.36$, $SD = 2.52$) compared to those who were presented with group-as-target threat ($M = 14.45$, $SD = 2.04$) or the control condition ($M = 13.73$, $SD = 2.34$). Similarly, the perceptions of the self-as-target group about their L2 teachers' gender stereotypical beliefs ($M = 14.36$, $SD = 2.72$) did not significantly differ from the group-as-target condition ($M = 14.79$, $SD = 1.16$) and the control group ($M = 14.58$, $SD = 1.41$).

Even though the stereotype manipulation check was non-significant, it was decided that the analyses in the following sections were still conducted. It is possible that the threat manipulation was effective (i.e. it worked to influence gender stereotype beliefs), but it was not sensitive enough in the current study. This might be due to the stereotype manipulation was not task specific (i.e., it was not related to L2 listening specifically) whereby responses were about L2 learning more generally. Therefore, the below analyses in Study 3 should be read with this caveat in mind.

Table 6.4. Univariate Test Results – Stereotype Manipulation Check

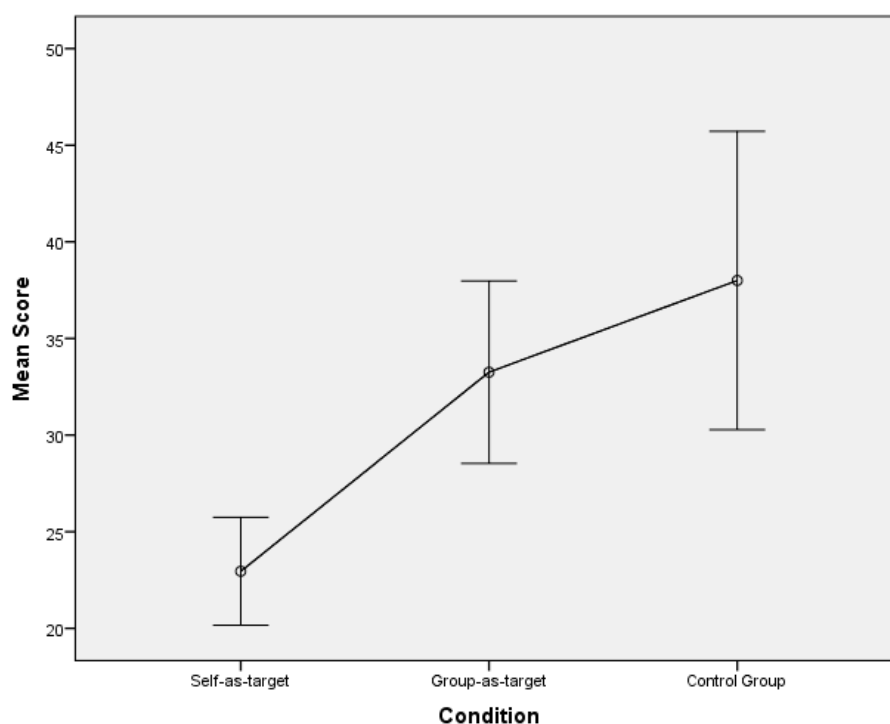
	<i>F</i>	<i>df</i>	<i>p</i>	η_p^2
Learner Beliefs	.812	2, 95	.447	.017
Teacher Beliefs	.416	2, 95	.661	.009

a. Computed using alpha= .05

6.5.4. English Listening Test Performance

A series of one-way Analysis of Variance (ANOVA) were performed to determine whether there were any statistically significant differences between the mean scores of the English listening test and its sub-sections across the three conditions (i.e., self-as-target, group-as-target and control conditions).

Figure 6.1. One-way ANOVA results for the English Listening Test Scores across the groups



First, participants' overall scores were compared. To determine the effect size, Cohen's (1988) guidelines were followed (i.e., 0.01 = small, 0.06 = medium, and 0.14 = large). Based on these guidelines, the results suggest that there was a significant main effect of the condition on the performance: $F(2, 100) = 5.15$, $p < .01$, $\eta_p^2 = .17$. The pairwise

comparisons indicated that participants in the self-as-target group had lower L2 listening scores ($M = 22.95$, $SD = 8.96$) relative to the group-as-target condition ($M = 33.26$, $SD = 13.74$) and the control condition ($M = 38.00$, $SD = 19.10$). There was a significant difference between self-as-target stereotype threat condition and the control condition, ($p < .05$). The difference between the self-as-target and group-as-target conditions was also significant ($p < .05$). However, there was not a significant difference between the group-as-target condition and the control group ($p > .05$).

As for the sub-sections of the English listening test, experimental condition did not have a significant effect on Part 4: $F(2, 100) = 1.92$, $p > .05$, $\eta_p^2 = .08$. However, there was a significant effect of the condition on Part 1: $F(2, 100) = 4.06$, $p < .05$, $\eta_p^2 = .11$, part 2: $F(2, 100) = 4.34$, $p < .05$, $\eta_p^2 = .06$ and Part 3: $F(2, 100) = 3.22$, $p < .05$, $\eta_p^2 = .07$. In Part 1, the pairwise comparisons revealed that the participants in the self-as-target group had lower scores ($M = 1.55$, $SD = .99$) relative to the group-as-target condition ($M = 2.26$, $SD = 1.61$) and the control condition ($M = 2.73$, $SD = 1.40$). Self-as-target stereotype threat had a significant effect on performance in comparison to the control condition, ($p < .01$). The difference between the self-as-target and group-as-target conditions, however, was not significant ($p > .05$). A similar pattern was observed in Part 2 in terms of the test scores. The participants in the self-as-target group had fewer correct answers ($M = 1.00$, $SD = 1.05$) relative to the group-as-target condition ($M = 1.69$, $SD = 1.27$) and the control condition ($M = 1.73$, $SD = 1.71$). However, the differences between the groups was not significant ($p > .05$). Similarly, the scores of the participants in the self-as-target group were lower in Part 3 ($M = 2.05$, $SD = 1.26$) relative to the group-as-target condition ($M = 2.74$, $SD = 1.48$) and the control condition ($M = 2.96$, $SD = 1.14$). When the pairwise comparisons were analysed, a significant difference was observed between the self-as-target and the control group ($p < .05$). The other pairs did not significantly differ from each other (all $ps > .05$).

6.5.5. L2 Reception Self-Efficacy and L2 Listening Anxiety

The participants' L2 reception self-efficacy and L2 listening anxiety were also analysed using a one-way Analysis of Variance (ANOVA). Table 6.5 presents the descriptive statistics for L2 reception self-efficacy and L2 listening anxiety along with L2 performance. The results indicated that there was not a significant main effect of the condition on the participants' L2 reception self-efficacy: $F(2, 98) = .73$, $p > .05$, $\eta_p^2 = .03$ and L2 listening anxiety: $F(2, 98) = .47$, $p > .05$, $\eta_p^2 = .006$. As mentioned earlier, such a non-significant result suggests that there is no need to provide any further tests. However, the following pairwise comparisons have been presented in the current thesis

for exploratory purposes only. The results gained from the pairwise comparisons showed that the participants in the self-as-target condition had the lowest L2 reception self-efficacy ($M = 13.00$, $SD = 2.76$) compared to the control group ($M = 14.04$, $SD = 3.16$) and the group-as-target condition ($M = 14.24$, $SD = 2.84$). While the mean differences between the self-as-target condition and the control group was not significant ($p > .05$), they significantly differed between the self-as-target condition and the group-as-target condition ($p < .05$).

Interestingly, the participants in the group-as-target condition had a higher level of L2 reception self-efficacy relative to the control group which suggests that the participants in the group-as-target condition might be motivated to disconfirm the negative stereotype attributed to their gender group. However, the difference between the two groups was not statistically significant ($p > .05$). L2 listening anxiety results were also surprising in that the participants in the control group had the highest mean ($M = 14.69$, $SD = 4.76$) relative to the experimental conditions, self-as-target ($M = 14.14$, $SD = 4.15$) and group-as-target ($M = 13.85$, $SD = 4.33$). Consistent with the L2 reception self-efficacy results, the participants in the group-as-target condition had the lowest L2 anxiety mean suggesting that they tended to disconfirm the gender stereotype presented to them. However, the mean differences between the groups was not statistically significant as in L2 reception self-efficacy ($p > .05$).

Table 6.5. Descriptive statistics for L2 self-efficacy, L2 listening anxiety and L2 performance

	Self-as-target	Group-as-target	Control
	Mean (SD)		
L2 Reception Self-efficacy	13.00 (2.76)	14.24 (2.84)	14.04 (3.61)
L2 Listening Anxiety	14.14 (4.15)	13.85 (4.33)	14.69 (4.76)
L2 performance	22.95 (8.96)	33.26 (13.74)	38.00 (15.00)

6.5.6. The Mediation Analysis

Following the initial analyses, a parallel multi-mediation model was built and run in SPSS using PROCESS macro 3.1v by Andrew F Hayes (Hayes, 2013) to establish the extent to which the effects of stereotype threat on L2 listening performance were mediated by the participants' L2 reception self-efficacy and L2 listening anxiety (Figure 6.2).

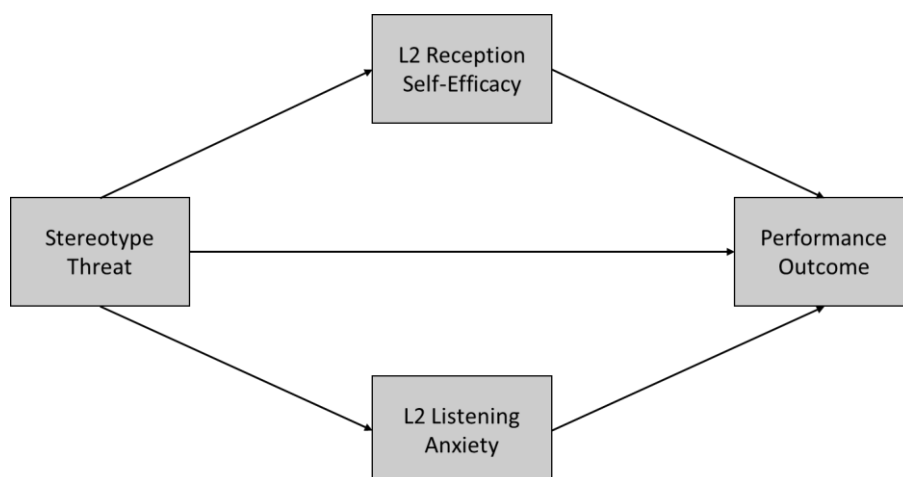


Figure 6.2. *The Mediation Model*

Given that the independent variable (i.e., the stereotype threat condition) in the current study was categorical, $k - 1$ independent dummy variables ($k=3$) were created (Hayes & Preacher, 2014). The self-as-target condition was selected as a reference group as it was the one with the numerically lowest value. Figure 6.3 presents the comparison results gained from the mediation analysis.

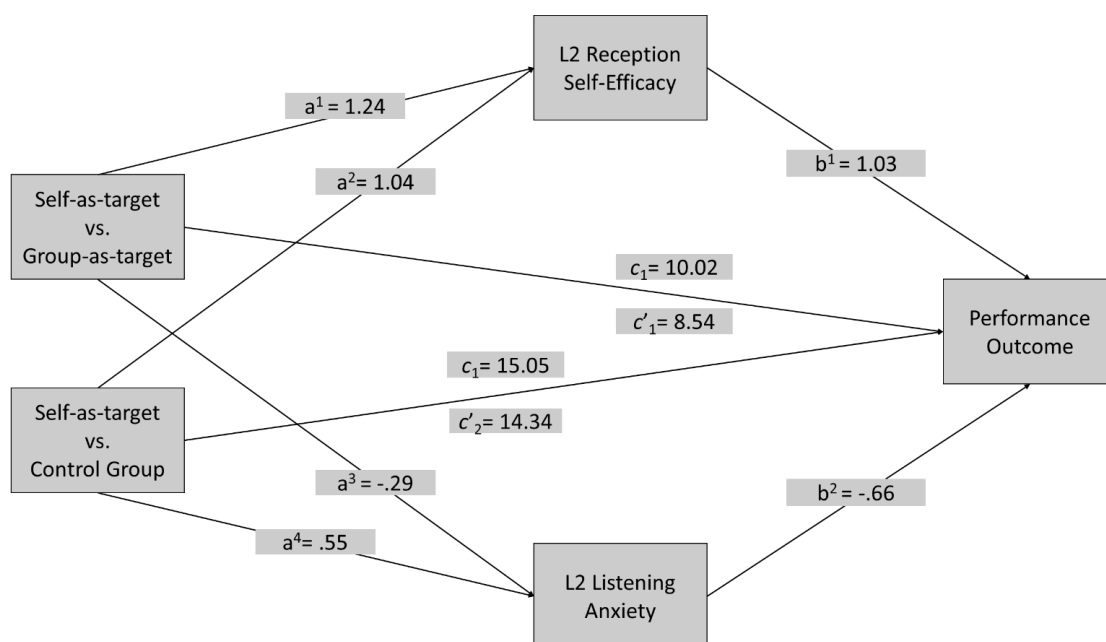


Figure 6.3. *The Mediation Model - The difference between the self-as-target and group-as-target conditions*

Based on 5,000 bootstrap samples, the self-as-target group was compared to the group-as-target and control groups. When the self-as-target group was compared to the group-as-target group, the results showed that the total effect (c_1) of the stereotype threat conditions on the performance was significant. The path analysis demonstrated that both L2 reception self-efficacy and L2 listening anxiety predicted the performance as the confidence intervals did not cross zero. However, neither L2 reception self-efficacy nor L2 listening were significantly associated with the stereotype threat conditions. As such, the indirect paths from the stereotype conditions to the performance through L2 reception self-efficacy and L2 listening anxiety were not statistically significant indicating that L2 reception self-efficacy and L2 listening anxiety did not mediate the link between the stereotype threat conditions and the L2 performance (Table 6.6).

Table 6.6. Direct, Indirect and Total Effects and 95% Confidence Intervals for the Mediation Model – Self-as-target vs. Group-as-target

		95% CI			<i>p</i>
		<i>B</i>	Lower bound	Upper bound	
Path	Direct Effects				
a^1	ST → SE	1.24	-.10	2.58	$p > .05$
a^3	ST → ANX	-.29	-2.32	1.73	$p > .05$
b^2	ANX → PERF	-.66	-1.26	-.05	$p < .05$
b^1	SE → PERF	1.03	.11	1.95	$p < .05$
c'_1	ST → PERF	8.54	2.46	14.62	$p < .05$
Indirect effects					
$a^1 + b^1$	ST → SE → PERF	1.28	-.22	3.54	$p > .05$
$a^3 + b^2$	ST → ANX → PERF	.09	-1.23	1.84	$p > .05$
Total effect					
$a^*b + c'_1$	ST → PERF	10.02	3.73	16.30	$p < .05$

Note: ST = Stereotype Threat, SE = L2 Listening Self-efficacy, ANX = L2 Listening Anxiety, PERF = L2 Performance

When the self-as-target group was compared to the control group, a similar pattern was observed for the direct and indirect paths. The mediation analysis revealed that the total effect (c_2) of the stereotype threat conditions on the performance was significant. The path analysis demonstrated that although both L2 reception self-efficacy and L2 listening anxiety predicted the performance, the stereotype threat conditions did not have a significant impact on either L2 reception self-efficacy or L2 listening self-efficacy. Accordingly, the indirect paths from the stereotype conditions to the performance through L2 reception self-efficacy and L2 listening anxiety were not statistically significant. This showed that neither L2 reception self-efficacy nor L2 listening anxiety was a mediator of the link between the stereotype threat conditions and the L2 performance (Table 6.7).

Table 6.7. Direct, Indirect and Total Effects and 95% Confidence Intervals for the Mediation Model – Self-as-target vs. Control group

		95% CI			<i>p</i>
		<i>B</i>	Lower bound	Upper bound	
Path	Direct Effects				
a^2	ST → SE	1.04	-.40	2.47	$p > .05$
a^4	ST → ANX	.55	-1.62	2.72	$p > .05$
b^2	ANX → PERF	-.66	-1.26	-.05	$p < .05$
b^1	SE → PERF	1.03	.11	1.95	$p < .05$
c'_2	ST → PERF	8.54	2.46	14.62	$p < .05$
Indirect effects					
$a^2 + b^1$	ST → SE → PERF	1.07	-.33	3.68	$p > .05$
$a^4 + b^2$	ST → ANX → PERF	-.36	-2.23	1.28	$p > .05$
Total effect					
$a*b + c'_2$	ST → PERF	15.05	8.31	21.29	$p < .05$

Note: ST = Stereotype Threat, SE = L2 Listening Self-efficacy, ANX = L2 Listening Anxiety, PERF = L2 Performance

6.6. Discussion

A large and growing body of literature has investigated stereotype threat, individuals' tendency to confirm negative stereotypes when they are explicitly activated, in a wide variety of areas. However, there is little empirical evidence as to whether this phenomenon occurs in foreign language learning contexts as well. The aim of the current study was therefore to contribute to the growing area of research by addressing this gap in the literature. This study examined the extent to which gender stereotype threat (i.e., that language learning is a female domain) had an impact on male L2 learners' language performance through the mediating roles of L2 reception self-efficacy and L2 listening anxiety. Underpinned by the multi-threat framework (Shapiro & Neuberg, 2007), it particularly focused on the difference between self-as-target and group-as-target stereotype threat and aimed to demonstrate whether the strength of the impact of these primes differed from each other. It further investigated the link between the learners' stereotype endorsement level and their susceptibility to gender stereotype threat effects.

The findings indicated that the participants in the self-as-target condition had lower performance scores compared to the participants in the group-as-target condition and the control group. As expected, the control group got the highest scores in the English listening test. However, the difference between the control group and the group-as-target did not significantly differ from each other. These findings suggest that gender stereotype threat might have an impact on male L2 learners' language performance, and consistent with the stereotype threat literature, the stereotype threat effect is greater when stereotype threat targeted the individual learners, rather than the whole gender group (Pennington, 2016). This finding should be evaluated considering Shapiro and Neuberg's (2007) profile of eliciting conditions necessary to yield each of the proposed stereotype threats. According to this profile, for people to experience self-as-target threat, they need to care about the implications of their stereotype-relevant actions for the way they see themselves. Similarly, for people to experience group-as-target threat, they need to care about the implications of their stereotype-relevant actions for the way they see the group. In the context of L2 learning at university level which was the focus of the current study, it is highly likely that the participants cared more about their own performance than their group's performance since any failure in the test would mean that the participants were not competent enough in English. Consequently, they would not be able to pass the English preparatory class which was the prerequisite for their further education at university. As the consequence of the language learners' poor performance would be more detrimental to themselves rather than the group in the chosen context, the

participants in the self-as-target condition performed worse than the one in the group-as-target condition.

In the current study, L2 reception self-efficacy and L2 listening anxiety were not found to be the mediators of the link between the stereotype threat conditions and language performance. This is in contrast with the results presented by Chung et al. (2010) who found that the state anxiety and self-efficacy mediated the stereotype threat effects African American's promotional exam performance. Unlike this study, self-efficacy and anxiety were not the mechanisms through which the stereotype threat effect was transmitted. These results can be explained in line with the multi-threat framework which suggests that distinct experiences of stereotype threat may be mediated or moderated by different factors (Shapiro & Neuberg, 2007). For example, previous research has shown that group identification might moderate the effects of stereotype threat based on the type of stereotype threat. In their study, Wout et al. (2008) assessed women and their maths performance and found that women who highly associated themselves with their gender had low performance when they were primed with a group-as-target stereotype threat. However, in self-as-target stereotype threat conditions, both women with high and low gender identification underperformed. It could therefore be argued that in the current study, some other factors such as gender identification might have played a role in the effects of stereotype threat on the learners' performance through L2 reception self-efficacy and L2 listening anxiety.

An alternative explanation to non-significant results gained from the mediation analyses is that individuals who are under stereotype threat might be motivated to disconfirm the stereotype attributed to themselves or their group (Jamieson & Harkins, 2007). In the current study, this trend was observed among the participants in the group-as-target condition. They had a higher level of L2 reception self-efficacy relative to the participants in the control group. Also, their L2 listening anxiety was lower than the control group suggesting that the participants in the group-as-target group were motivated to show that the given gender stereotype could not be applied to them. It may also be the case that they put extra effort in the English listening test to disconfirm the gender stereotype because their performance did not significantly differ from the control group. Such a result was not observed among the participants in the self-as-target condition or it was not as significant as it was in the group-as-target condition which again highlights the importance of differentiating between the distinct stereotype threats. It is important to reiterate here the fact that the stereotype manipulation check was non-significant in the current study. These results therefore need to be interpreted with caution.

6.6.1. Limitations and Suggestions for Future Research

It is important to note that there are a number of potential limitations that should be considered when interpreting the findings of this study. In the current research, the focus was on the target of stereotype threat (i.e., whether individuals' stereotypical behaviours reflect upon the self or the group). It was explored whether L2 learners' performance was significantly debilitated by self-as-target or group-as-target threat. However, the multi-threat framework has also another important dimension, namely the source of stereotype threat (i.e., whether individuals' performance will be judged by out-group members or in-group members) that needs to be investigated. In the present study, the participants were not explicitly exposed to the source of stereotype threat. However, the participants might have been concerned that their performance scores would be judged by ingroup or outgroup members such as their language teachers. Therefore, their L2 listening performance might be impacted by both the target and the source of stereotype threat. Future research should determine whether the source of stereotype threat results in performance decrements as well.

Also, the current research investigated the roles of two mediators (i.e., L2 reception self-efficacy and L2 listening anxiety) and one moderator (i.e., stereotype endorsement) in the stereotype threat effects on the participants' L2 performance. However, there are a number of other mediators and moderators that might explain the link between stereotype threat and performance. For example, previous research has documented that domain identification (i.e., the importance of the domain tested) might be a prerequisite of stereotype threat (Pennington et al., 2016). It is, therefore, suggested that future studies should take such factors into account and investigate which mediators and moderators are more likely to heighten the effects of stereotype threat. As the distinct types of stereotype threat might be mediated or moderated by different processes, more research is needed to better understand the association between the types of stereotype threat and the different mediators and moderators.

Another limitation of this study was concerned with the task chosen. Due to the practical constraints (i.e., easier to administer and score), the current research focused only on L2 listening which is one of the most anxiety provoking skills. However, there are three more language skills (i.e., reading, writing and speaking) that need to be studied in the context of stereotype threat. It is possible that some language skills might make L2 learners more susceptible to stereotype threat. Therefore, a natural progression of this work could be to assess the effects of stereotype threat in the other three language skills,

especially speaking which is believed to provoke the highest levels of anxiety for FL learners. The need for a special focus on listening and speaking skills was evidenced by the interviews conducted in Study 2 as well. More than half of the learner participants stated that their past EFL learning experiences indicated that EFL teaching and learning had mostly depended on memorisation of grammar and vocabulary. The lack of meaningful learning resulted in their incompetence in the communication skills, namely listening and speaking. Since they did not have enough opportunities to practice these skills adequately prior to university, their self-efficacy might be lower and anxiety level might be high. Therefore, further studies, which take these variables into account, will need to be undertaken.

Also, like Study 1, the one sample *t*-tests revealed that English was not as gendered as some of the academic domains tested in the current study (please see Section 6.5.2 for the one-sample *t*-tests). Therefore, it is possible that the belief that English is a female domain was not strong enough among the participants to create a significant stereotype threat effect, which might have resulted in the non-significant link between the gender stereotyped beliefs and the performance via L2 anxiety and L2 self-efficacy (for males only).

6.6.2. Original Contribution to Knowledge

Taken together, the current research makes three noteworthy contributions to the stereotype threat literature. First, although there is an extensive literature on stereotype threat, most of them focus on the academic subjects such as science and math. Therefore, this study represents one of few empirical studies that have investigated the stereotype threat phenomenon in the field of foreign language learning. Not only did it delve into the multi-threat framework by looking at the distinct experiences of stereotype threat (i.e., self-as-target and group-as-target), it also explored how stereotype threat manifests itself through the mediating roles of L2 reception self-efficacy and L2 listening anxiety.

Secondly, many studies focus on female learners and their achievement in the academic subjects which are believed to be a male domain such as maths and science. In this study, it was proposed that it is not only female learners who are in danger of underperforming in certain academic subjects. There might also be a threat for male learners who are in the academic subjects widely known as a female domain such as foreign language learning. This study suggested that stereotype threat might be an issue

for male learners as well in a female oriented academic field as they showed poor performance in English listening test under the stereotype threat conditions.

Finally, as noted by Owens and Massey (2011), experiments conducted in labs confirm the hypothesis of stereotype threat with a high degree of internal validity. However, such studies do not have external validity since results cannot readily be generalized to real-world settings such as school or university settings. The participants in this study were tested in a real-world context to their learning (i.e., in real language classrooms) meaning that domain identification was likely to be more controlled than in much other ST research which generally takes place in labs. Therefore, a strength of this research was that it was in a more ecologically valid context than much other ST research.

6.7. Chapter Summary

Chapter 6 has presented the final empirical study of this thesis which experimentally tested whether self-as-target or group-as-target threats could independently impair male EFL learners' English listening performance. Also, it was tested whether L2 listening anxiety and L2 reception self-efficacy mediated the effect of these threats on English listening performance. It was found that compared to the participants in the group-as-target condition, the ones in the self-as-target condition performance worse in English speaking test. However, neither L2 listening anxiety not L2 reception self-efficacy did not mediate the effects of stereotype threat in any conditions. Additionally, it was examined whether stereotype endorsement level moderated the effects of the mediators. The findings demonstrated that stereotype endorsement did not moderate the mediating effects of L2 listening anxiety and L2 reception self-efficacy. The results of this study are highlighted further in the next chapter, Chapter 7 which integrates the findings of all the three studies and discuss them with references to the research literature and research questions framing this thesis.

CHAPTER 7: GENERAL DISCUSSION

7.0. Chapter Outline

Chapter 7 sets out to present the general discussion and the conclusion of the thesis. The chapter is divided into three main sections. The first section provides an overview of the findings of Study 1, Study 2 and Study 3. It also aims to integrate the findings and draw inferences based on the mixed methods findings. The second section discusses the limitations of the current research and highlights the scope for future research. Finally, the last section is concerned with the original theoretical and practical contributions of this work to the wider literature. It further presents the implications for practice followed by the concluding remarks.

7.1. Summary and Integration of Findings

As shown in Figure 7.1, the current research was based on the hypothesis that L2 performance is linked to gender stereotyping of L2 learning via L2 anxiety and L2 self-efficacy. This hypothesis was explored through three different, but related studies. Overall, the studies aimed to address the following research questions:

1. Is there a relationship between L2 learners' existing gender stereotyped beliefs about L2 learning and their L2 self-efficacy, anxiety and performance?
2. To what extent does the learning environment foster the salience of gender-stereotyped beliefs?
3. What is the impact of making gender-stereotyped beliefs salient on L2 learners' L2 self-efficacy, anxiety and performance?

A mixed methodology research design was adopted to better explore the aforementioned research questions. The rationale for using a mixed-methods approach was described in detail in Chapter 3. Briefly, mixing the quantitative and qualitative methods was ideal for this work for two main reasons. First, it could address the variety of my research questions provided above. As Cohen et al. (2013) state, MMR design enables researchers to study both the 'what' (numerical and qualitative data) and 'how or why' (qualitative) types of research questions. Secondly, it provides a bigger picture. Johnson and Christensen (2000) emphasise that quantitative and qualitative research methods are complementary. Using a single approach might be inconclusive because all research approaches have different strengths and weaknesses. Therefore, by combining the approaches, I aimed to achieve a more complete understanding of the phenomenon of gender stereotyping of L2 learning.

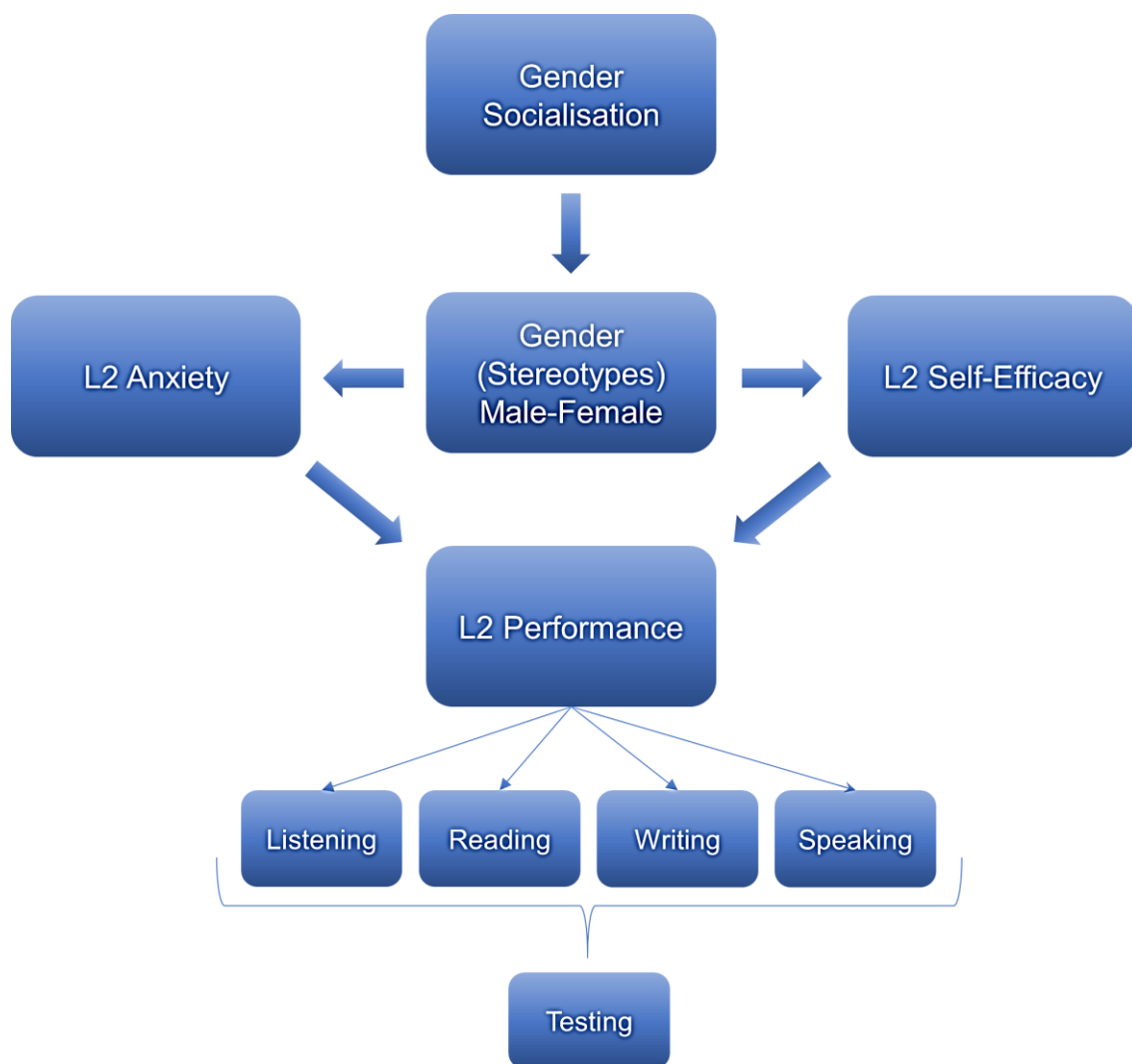


Figure 7.1. *The Hypothesis of The Current Research*

From among four main mixed methods designs, the convergent parallel design was chosen in the current study (see Figure 7.2). This design allows researchers to acquire different, but complementary data on the same phenomenon. As the nature of the design suggests, the quantitative and qualitative strands of this study were implemented during the same phase of the research. All the methods had an equal weight at the time of data collection, and the strands were treated independently during the analysis stage (see Chapter 4, 5 and 6). As seen in Figure 7.2, the current chapter is designed to integrate all the findings and provide an overall interpretation.

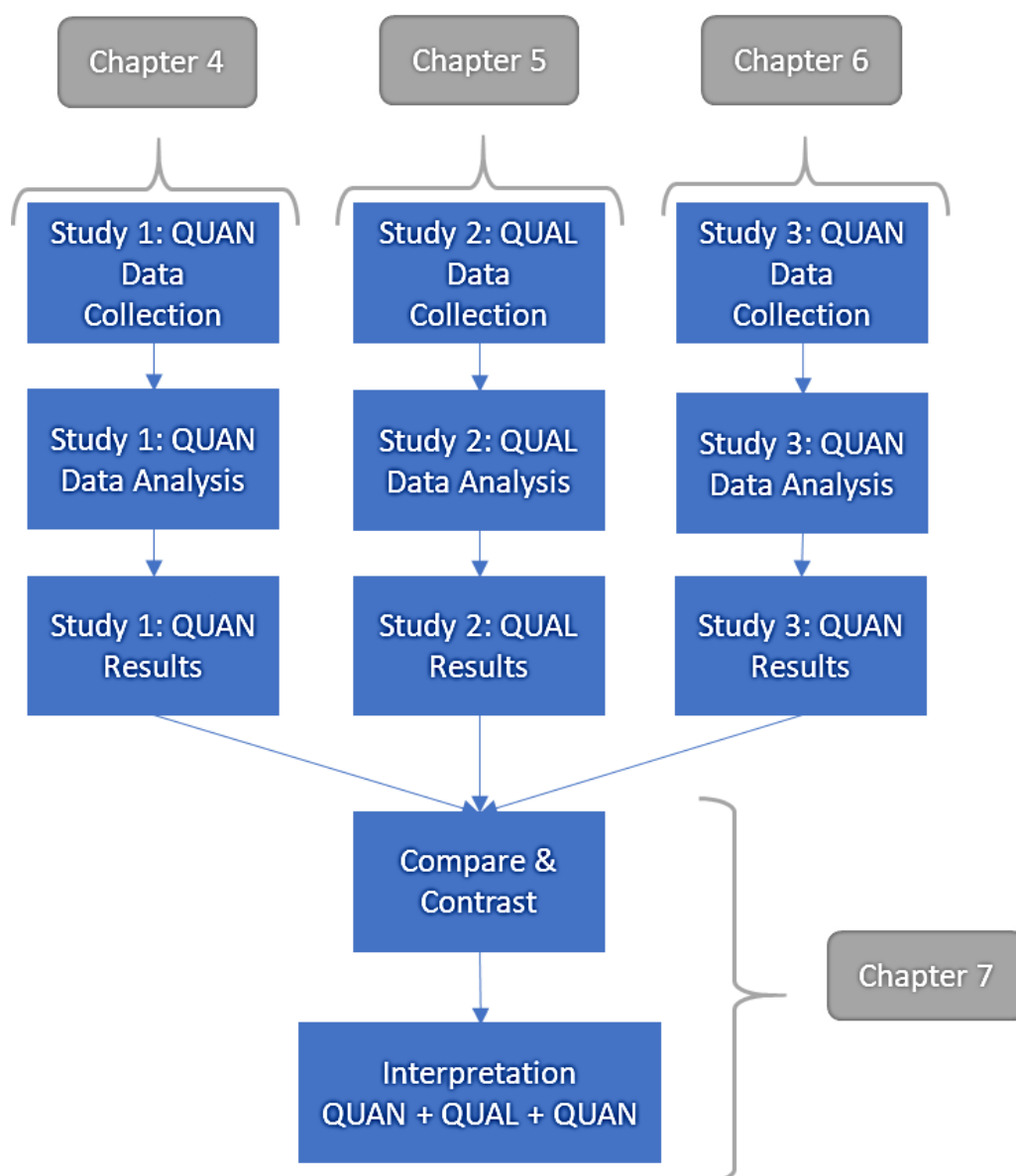


Figure 7.2. *The Convergent Parallel Design of the Current Research*

The research approach itself was twofold. In the first and second phases of the research, a naturalistic approach was adopted. In the first phase, the participants' gender-subject stereotyped beliefs across a number of curriculum subjects, as well as specifically to language learning were examined. This aimed to assess the extent to which these perceptions were related to self-report measures of L2 anxiety, L2 self-efficacy and L2 performance. In the second phase of the research, the role of teachers in sustaining or legitimising these gender stereotyped beliefs were explored in depth with the help of interviews with L2 teachers and learners. In the third phase, an experimental approach was adopted. It was investigated whether the phenomenon called stereotype threat, the tendency to confirm gender stereotypes when they were explicitly activated also occurred in the language learning context. Each phase was explored in detail in the

previous chapters (please see chapter 4 for Study 1, Chapter 5 for Study 2 and Chapter 6 for Study 3). The next sections integrate the findings gained from all these three studies and discuss the results drawing upon the relevant literature.

7.1.1. Gender Socialisation and Gender Stereotyping of L2 learning

As seen in Figure 7.1, it was crucial in the current research to determine the extent to which the commonly acknowledged gender stereotypes pertaining to L2 learning was accepted among Turkish adult learners of English. Study 1 demonstrated that consistent with the previous literature (Carr & Pauwels, 2006; Pomerantz, 2008; Schmenk, 2004), the gender stereotype that females are better at L2 learning was significantly endorsed by both females and males. These findings were supported by the findings gained from Study 2. Even though the participants did not explicitly believe that L2 learning was a 'feminine' domain, most of the participants agreed that there were significant gender differences in dispositional attributes (e.g., perseverance and sense of responsibility). The results indicated that females showed greater perseverance in the face of difficulty while learning a new language compared to males. Also, their sense of responsibility was higher than males (e.g., doing homework). Additionally, the findings gained from Study 2 showed that there were certain differences in showing emotions as well. It was revealed that L2 male students did not tend to show their emotions such as anxiety easily compared to females.

Although it was emphasised during the interviews with both the L2 learner and teacher participants that both females and males can be successful language learners, the aforementioned findings suggest that females might be more advantageous in L2 learning due their attitudes towards studying. Also, females' being comfortable with their emotions might help them get more attention from their language teachers compared to male L2 learners. Even if such differences would not matter for advanced learners or for those (female or male) who have learning disposition, the ability to work hard, show determination or showing emotions such as anxiety might play a crucial role in the beginner levels. Since it was beyond the scope of this research to evaluate the link between the dispositional attributes and different language levels, this is further examined in the limitations and suggestions for future work section below in Section 7.1.3.

With regards to the participants' beliefs about their teachers' perceptions of L2 learning, in Study 1, female learners endorsed that their teachers believed more females than males were better at L2 learning. Similarly, male learners also believed that their

teachers had gender stereotypical beliefs about L2 learning. Expanding on this, Study 2 explored whether L2 teachers expressed any gender stereotypical beliefs themselves, and the extent to which their beliefs played any sustaining and legitimising role in L2 learners' gender stereotyped beliefs, if any (RQ 1). It was revealed that the L2 teacher participants had some gender stereotyped beliefs pertaining to L2 learning which was in line with the results gained from the learner participants in Study 1. However, unlike expectations, Study 2 showed that L2 teachers did not play a significant role in either sustaining and legitimising L2 learners' own gender stereotyped beliefs. That is, even if the L2 teacher participants thought that females might have more sense of responsibility or they showed greater perseverance in the face of difficulty, these beliefs did not reflect on their attitudes and behaviours in or outside the classroom. This was confirmed through the interviews with L2 learners as well. Almost all L2 learner participants indicated that their teachers did not differentiate between males and females and treat them equally (RQ2).

Since it was revealed that it was not L2 teachers at university who created, sustained or legitimised the gender stereotypes held by L2 learners, it was important to explore the other possible sources of the existing gender stereotypes pertaining to L2 learning. This was not possible in Study 1 which asked the learner participants to indicate whether they agreed with the common gender stereotypes pertaining to L2 learning. That is, there was not an opportunity for the participants to expand on their responses or explain the possible sources of any existing gender stereotypes they might held. To address this limitation, in Study 2, both L2 learner and teacher participants were asked to share their thoughts and beliefs about the sources of their gender stereotypical beliefs. One of the important findings in Study 2 was that most of the existing gender stereotyped beliefs which the L2 learner participants had were formed prior to university. Some of the L2 learner and teacher participants mentioned that they had some teachers at primary or secondary schools who had gender stereotypical behaviours (e.g., those who believed that females were better). This is also further scrutinized in Section 7.1.3.

Study 2 showed that the social norms and expectations were found to be the predominant source of the aforementioned gender differences. According to some of the participants, the gender differences mainly stemmed from the different social expectations of females and males in Turkey. While males were always expected to be strong and fearless, females were seen as fragile and in need of protection. Several participants also mentioned that when males felt anxious, they had different strategies

to protect their self-image including making jokes or being completely silent in the classroom.

One of the interesting findings was that the majority of the participants agreed that females and males might have different career aspirations, and this might create some gender imbalances in the field of languages. Most of the participants mentioned that they had more female language teachers than males. This was also believed to be due to certain social expectations of females and males in terms of a career path. While females were expected to do jobs that they could easily perform along with their roles as a spouse and mother, males were expected to do jobs that they would earn enough money as a 'breadwinner'. As the English related occupations were not seen in the latter category, males did not generally prefer pursuing a career in English. As such, L2 learning might be mostly associated with females rather than males. These results were in line with van der Vleuten (2016) who suggested that if females and males would like to conform to their gender ideology, it is highly likely that they would prefer more masculine and feminine subjects, respectively.

Differently from Study 1 and Study 2, Study 3 tested the impact of making gender-stereotyped beliefs salient on Turkish male L2 learners' gender stereotype endorsement level. Surprisingly, there were not any significant differences between the control (i.e., no threat) and experiment groups (i.e., those primed with the stereotype threat that 'Females are better than males in L2 learning') in terms of the stereotype endorsement level. The results suggested that the group presented with the self-as-target threat did not hold significantly different gender stereotypical beliefs concerning L2 learning compared to those who were presented with group-as-target threat or the control condition. Similarly, the perceptions of the self-as-target group about their L2 teachers' gender stereotypical beliefs did not significantly differ from the group-as-target condition and the control group (RQ3). When the means of the groups were compared, it was seen that the male participants tended to disconfirm the gender stereotypes when they were made explicit in the experimental groups. These results match those observed in Study 2 which suggested that Turkish male L2 learners tended to protect their self-image as a strong and confident person. They also corroborate the ideas of Schmader (2002) who suggested that those who highly identify themselves with the group feel more pressure to disconfirm the negative stereotype in order not to be a bad ambassador of the group. This result may be explained by the fact that Turkish male learners identify themselves with their gender and this could reflect on their behaviours and attitudes while learning a new language.

7.1.2. The Mediating Role of L2 Anxiety and L2 Self-efficacy

As shown in Figure 7.1, the hypothesis was that gender stereotyped beliefs were linked to L2 performance via two mediators, L2 anxiety and L2 self-efficacy. It was expected that the gender stereotype belief that females were better than males in L2 learning would increase female L2 learners' self-efficacy and decrease their L2 anxiety while the link between gender stereotypes and the mediators would be the other way around for males. That is, the more they believed females were better, the more were they expected to have high L2 anxiety and low L2 self-efficacy. Accordingly, it was assumed that male L2 learners who held the belief that females were better would score lower in the final L2 exam compared to females due to their increased L2 anxiety and decreased L2 self-efficacy. The participants' perceptions of their teachers' gender stereotyped beliefs were also tested based on this hypothesis.

In Study 1, the findings confirmed that females who believed that females were better in L2 learning showed a better performance in L2 exam via an increased L2 self-efficacy. However, the link between their gender stereotyped beliefs and L2 performance was not mediated by L2 anxiety. It is highly likely that since L2 anxiety and L2 self-efficacy were strongly correlated, it could not be specified the extent to which L2 anxiety and L2 self-efficacy independently mediated the link between the gender stereotyped beliefs and L2 performance. The results concerning L2 female learners' perceptions of their teachers' beliefs were along the similar lines with the aforementioned findings. The female participants who perceived their teachers to hold a belief that L2 learning was a male domain performed worse due to lower L2 self-efficacy. However, L2 anxiety did not mediate the link between the gender stereotyped beliefs and L2 performance.

As for L2 male learners, neither L2 self-efficacy nor L2 anxiety were found to be the mediators of the link between the gender stereotyped beliefs and their L2 performance (RQ1). Similarly, no significant link was found between L2 male learners' perceptions of their language teachers and their performance when mediated by L2 anxiety and L2 self-efficacy. It is possible that the non-significant results were due to the shared variance between L2 anxiety and L2 self-efficacy. However, there are also some other possible explanation for the results.

As mentioned above, Study 2 revealed that there are some certain social norms and expectations of males and females in Turkey. One of the findings was that Turkish male learners, especially those who identify themselves with their gender, did not tend to show

their emotions such as anxiety and attempted to be strong and confident compared to females for whom showing emotions or being fragile was more likely to be accepted by the society. Considering these findings, it is likely that in Study 1, some of the male L2 learners who were asked to share their perceptions about their L2 anxiety and L2 self-efficacy in Study 1 acted in accordance with the social expectations of males. That is, it is likely that they preferred presenting themselves as less anxious and with low self-efficacy in the questionnaires which they completed in Study 1.

A similar pattern was observed in Study 3 as well which examined the extent to the phenomenon of stereotype threat occurred in the context of L2 learning. It was found that the male participants in the group-as-target condition had a higher level of L2 reception self-efficacy relative to the participants in the control group. Also, their L2 listening anxiety was lower than the control group suggesting that the participants in the group-as-target group were motivated to show that the given gender stereotype could not be applied to them. It may also be the case that they put extra effort in the English listening test to disconfirm the gender stereotype because their performance did not significantly differ from the control group (RQ3). Such a result was not observed among the participants in the self-as-target condition or it was not as significant as it was in the group-as-target condition which again highlights the importance of differentiating between the distinct stereotype threats. This tendency among the male L2 learners are consisted with Jamieson and Harkins (2007) who suggest that individuals who are under stereotype threat might be motivated to disconfirm the stereotype attributed to themselves or their group.

The non-significant results can also be explained in line with the multi-threat framework which suggests that distinct experiences of stereotype threat may be mediated or moderated by different factors (Shapiro & Neuberg, 2007). In the current study, it was tested the extent to which L2 reception self-efficacy and L2 listening anxiety mediated the effects of two target of the stereotype threats (the self and the group) on the L2 listening performance. These findings seem to be consistent with other research which found that self-reported anxiety was not a mediator of the link between stereotype threat and performance (e.g., Steele & Aronson, 1995). Similarly, self-efficacy did not mediate the effects of self-as-target threat on African American's cognitive ability (Mayer & Hanges, 2003) and group-as-target threat on women's mathematical performance (Spencer et al., 1999). It could therefore be argued that some other factors such as perseverance, motivation and sense of responsibility might have played a role in the effects of the target of the stereotype threats on the learners' L2 listening performance.

Since the literature on the effects of gender stereotypes on language learners' performance through L2 self-efficacy and L2 anxiety is limited, it is relatively difficult to discuss the results presented in the current research in relation with other studies. Therefore, more research is encouraged to better understand the role of gender stereotypes and the phenomenon of stereotype threat in L2 learning (please see Section 7.2 for a detailed discussion of suggestions for future research).

7.2. Thesis Limitations and Suggestions for Future Research

The generalisability of the results of the current research is subject to certain methodological and theoretical limitations which need to be acknowledged. These are discussed in detail below.

7.2.1. The Sample

Although the sample size of the current research was quite large, all the participants recruited in the three studies were Turkish. Therefore, it is not possible to ascertain the extent to which the findings presented here can be generalised to other language learning contexts. As such, more research with different groups of learners is needed to get further support for the generalisability of the results gained from the three studies outlined in the above sections. In particular, it is needed to study contexts differ from each other in terms of gender roles and expectations of females and males. This will promote diversity in research and help us develop a better understanding of the scope and nature of gender stereotyping of L2 learning.

Also, in the current research, all the three studies were concerned with English as a foreign language. Since English has become the global language, English language learners' motivation might be different from those learning other languages. Therefore, it is probable that there are differences between these learners in terms of L2 anxiety and L2 self-efficacy which are believed to play a mediating role in the link between gender stereotyping of L2 learning and L2 performance. Given that, in future investigations, it might be possible to use a different language such as German, French or Italian and investigate the issue of gender stereotyping of L2 learning in respect of these languages.

This study was concerned with EFL learning at university level, and the participants were generally aged 18-22. The results suggested that most of the gender stereotypical beliefs were formed prior to university and by primary or secondary school teachers as well as families. Since language education prior to university was not the main concern in the

current research, it was not possible to explore this issue in depth. However, this is an important issue for future research. There is abundant room for further progress in determining the extent to which language teachers or families play a role in sustaining and legitimising the commonly accepted gender stereotypes pertaining to academic subjects including L2 learning.

As highlighted earlier, the current research did not focus on the differences between the levels of English language proficiency (e.g., beginner learners of English or advanced learners of English). That is, it was beyond the scope of this study to investigate whether gender stereotyping of L2 learning occurred differently within or across different levels of language proficiency. Therefore, a future study investigating possible differences between the different levels of language proficiency would be very interesting. This will help us determine whether a group of learners might be more or less susceptible to the impact of gender stereotyping of L2 learning.

Finally, it is important to note that gender is treated as a binary construct in the current study. That is, the participants were categorized as either male or female. However, it needs to be acknowledged that there are considerable differences within each gender group. According to Social Identity Theory (see Chapter 2 for a detailed review), people do not automatically have a sense of belonging in the groups that surround them (Hogg, 2006). The idea of groupness occurs when people identify themselves with the groups. Therefore, the extent to which the gender stereotypes and gender stereotyping of academic subjects have an effect on a person's self-efficacy or anxiety in chosen academic domains may depend on their group identification level. As such, although there are similar patterns when it comes to gender stereotyping and its effects, the results cannot be generalised for all males or females. Therefore, further research should be done to investigate the phenomenon of gender stereotyping of L2 learning considering non-binary gender identities.

7.2.2. Data Collection Instruments

Due to the practicality and feasibility issues, the participants' overall scores which they received at the end of their nine-month English education was used as the performance measure in Study 1 rather than adopting a standardised performance measure. Since there were four different universities involved in the study, the extent to which the overall scores were comparable to each other was questionable. Although all the universities involved in this research assessed the participants' ability in all the four skills, namely listening, speaking, reading and writing, these universities probably differed from each

other in terms of curriculum delivery and types and content of assessment. For example, it is possible that a language learner whose overall score was 50 in A University could get a higher or lower overall score in B university due to different types of instruction and assessment. Adopting a standardised performance measure could facilitate a more objective comparison. Therefore, it is suggested that the association of gender stereotyping of L2 learning and L2 performance is investigated adopting a standardised performance measure in future studies.

Another limitation is concerned with the assessment of L2 self-efficacy using the new Questionnaire of Self-efficacy Beliefs in Learning a New Language (QSLL). As Dörnyei (2000) suggests self-beliefs are not static but fluctuate over time. Therefore, it is suggested that these beliefs are measured a number of times over time. However, due to the limited time frame and the difficulty to access the same participants in different times, L2 self-efficacy could be measured only once. As such, it is suggested that the link between gender stereotyping of L2 learning and L2 self-efficacy is investigated adopting a longitudinal study in which L2 self-efficacy is measured a number of times over the course of language education.

Furthermore, Council of Europe (2017) has recently published a provisional edition of the Companion Volume. This is intended as a complement to the Common European Framework of Reference for Languages (CEFR) which informed the QSLL. Based on the project pursued between 2014 to 2017, this new document offers an updated version of the CEFR descriptors (2001) as well as introducing the new descriptors for new areas (Council of Europe, 2017). According to this new document, mediation (including reactions to creative text/literature), online interaction, and plurilingual/pluricultural competence need to be treated as part of language proficiency to address the increasing linguistic and cultural diversity of the societies. Therefore, future research may wish to extend the QSLL's domains of interest by adding the mediation construct or enrich the content of the QSLL by benefiting from the new and updated CEFR descriptors.

7.3. Contribution to Knowledge

Despite the aforementioned limitations, the empirical studies conducted in the current thesis have made a several noteworthy contributions to knowledge. These are discussed separately in the next sections.

7.3.1. Gender Stereotyping of Academic Domains

The empirical findings in this study provide a new understanding of gender stereotyping of academic domains since it demonstrated, for the first time, that gender stereotyping of academic domains could be an issue for males in so-called feminine academic fields as well. Although there is a large and growing body of literature on the concept of gender stereotyping of academic domains, it seems that much of the current literature has paid particular attention to the academic subjects such as maths and science which are believed to be masculine domains (see Chapter 1 for an overview). The current research explored the topic putting the emphasis on language learning which is mostly associated with females and determined the extent to which such stereotypes were an issue for males. Believing that the impact of gender stereotyping of academic domains on male learners' performance was not necessarily a direct one, this study examined the link between gender stereotyping of language learning and L2 performance via the mediating factors namely self-efficacy and anxiety. This new approach to the issue of gender stereotyping of academic domains is among the key strengths of this study.

In particular, this research made a significant contribution to the ST literature by focusing on a real-world environment in which an under-represented sample was tested. As discussed in Chapter 6, most of previous ST research has been conducted in labs rather than in real-world settings. Unlike previous research, this research was conducted in language learners' actual settings, namely classrooms. Since such an approach is new to the ST literature, it provides an important theoretical contribution which could inform future research.

7.3.2. The MLCAS

The need for a new scale assessing L2 anxiety was discussed in detail in Chapter 2. To address this limitation, the current study produced a new Multidimensional Language Class Anxiety Scale (MLCAS). Given that, this study has three major contributions to L2 anxiety research. First, as Pavlenko (2013) states, psychological factors such as motivation, anxiety and willingness to communicate do not stand alone. These are continuously interacting with many other personal (e.g., age and aptitude), social (e.g., ethnicity and culture), and situational (e.g., classroom environment and peers) factors as well as with each other (MacIntyre, 2017). In line with these suggestions, the 2010s have seen the rapid development of a more contextualised way of studying L2 anxiety which is called dynamic approach. Researchers have showed increased interest in the interactions between L2 anxiety and other factors including self-efficacy and enjoyment (Dewaele, 2013; Dewaele & MacIntyre, 2014; MacIntyre & Serroul, 2015; Mills et al., 2006). Using a concise scale measuring L2 anxiety pertaining to all the language skills

and testing simultaneously will be of great support to researchers who embrace the dynamic approach as it makes it possible for these researchers to assess L2 anxiety without compromising the length of the scale. Also, language teachers who aim to assess their students' level of L2 anxiety efficiently will find the briefness of this scale very practical and feasible.

Secondly, unlike the items in many other scales in the L2 anxiety literature, the items which were adapted from the AEQ (Pekrun et al., 2005) offer a unique and contemporary approach to measuring emotions such as anxiety. The MLCAS differentiate between affective, cognitive and physiological components of anxiety enhancing our understanding of emotions in two ways. Theoretically, it offers a better understanding of the nature of L2 anxiety which L2 learners suffer from. Practically, it makes it possible for language teachers to help anxious L2 learners appropriately.

Lastly, as a further contribution, I developed a measure of L2 test anxiety which is intended to correspond as an additional sub-scale on the MLCAS. To date, a measure that brings all the skills and testing together to assess L2 anxiety is not available. Previous research has generally used general test anxiety scales (e.g., Sarason, 1978) rather than devising a new measure which is specific to L2 learning believing that it is not a construct that is specific to L2 anxiety but rather a general concept. In this study, it has been shown that test anxiety is indeed a part of L2 anxiety and it should be assessed along with the language skills. Therefore, it could be possible to differentiate whether learners are in fact anxious because of a particular language skill (listening, reading, writing, and speaking) or it is just the nature of testing that makes them worried.

It should be noted that the MLCAS which embodies five sub-scales has been developed and tested both as separate subscales and as a whole in the present study, researchers are encouraged to use the scale as appropriate to the aims of a particular study. That is, they can use the scale to measure overall L2 anxiety or choose specific sub-scales if they wish to focus on any component of the MLCAS discretely.

7.3.3. The QSLL

The current study also developed and validated a new L2 self-efficacy scale that covers receptive (i.e., reading, listening) and productive skills (i.e., speaking and writing). Both the pilot and the main studies indicated that the QSLL reflected a two-factor structure (i.e., reception and production) accounting for the correlated residual variance between the specific language skills. The new measure showed good internal consistency as

evidenced by the high acceptable Cronbach's alphas ($\alpha > .80$) which confirmed that the data gained using the QSLL was reliable. Overall, this study has three significant contributions to the L2 self-efficacy literature. First, as suggested by Wang et al. (2013), to create the initial items, I utilised the CEFR (2001) which is a widely-known and used framework in Europe and increasingly, in other countries. Therefore, it will be easy and straightforward for most of researchers and language teachers to use the QSLL as they will already be familiar with the content and structure of it. Also, as most of second language instruction is designed based on this framework, the QSLL will be applicable to most of language learning contexts. Second, while creating our initial items, I followed Bandura's (2006) widely accepted guidelines. For example, the QESE (Wang et al., 2013), the only available measure aiming to cover all the four language learning skills so far, does not manage to address the need for gradations of challenges which is suggested by Bandura (2006). The QSLL, however, provides a mixture of easy and difficult items which would fill in the gap in the literature. Third, the QSLL confirms that L2 self-efficacy is not unidimensional and acknowledges the differences between productive and receptive skills. This supports the necessity of assessing the language skills separately rather than considering L2 self-efficacy as a simple and straightforward concept.

This research is timely and important in that with the increasing importance of self-beliefs in language teaching and learning, there is a continuous need for established measurements that would enable researchers to collect valid and reliable data. Our study provided some preliminary evidence for the reliability and validity of the data gained using the new Questionnaire of Self-Efficacy Beliefs in Learning a Foreign Language which covers self-efficacy concerning the four major language learning skills categorised under two broad constructs: reception and production. Although the research on self-efficacy related to the language learning skills is somewhat scarce, the results of the limited literature reveals the importance of its impact in the understanding of language learners' self-efficacy about the skills and to what extent this affects their performance. Therefore, the QSLL evaluating students' beliefs concerning the four skills can represent a major contribution to the existing literature.

7.4. Thesis Conclusions

Using a convergent parallel mixed methods design, the present study aimed to investigate the extent to which the phenomenon of gender stereotyping of academic subjects occurred in the context of L2 learning among Turkish adult EFL learners in Turkey. To address this research aim, three different but related studies were designed.

Study 1 examined whether there was a link between language learners' gender stereotyped beliefs about foreign language learning and their self-efficacy, anxiety, and performance. To collect quantitative data, three new questionnaires (i.e., MLCAS, QSLL, QLGB) were created, and these questionnaires underwent a rigorous design and validation process. Study 2 qualitatively explored the extent to which language teachers, as an agent of socialisation, played a role in sustaining or legitimising any existing gender stereotyped beliefs. Semi-structured interviews were conducted with both L2 teachers and L2 learners to provide a fuller picture of the phenomenon. Finally, Study 3 experimentally investigated the impact of stereotype threat pertaining to learning another language upon male language learners' performance via their self-efficacy and anxiety. In particular, it was examined whether self-as-target or group-as-target threats had a different effect on the participants' English listening test.

The data collected from the three studies were analysed being informed by the paradigm of critical realism. The results in Study 1 suggested that both female and male EFL learners endorsed the belief that females were better EFL learners. However, this belief was not linked to their L2 performance via L2 anxiety. L2 self-efficacy was the mediator of this link only for females (RQ 1). In Study 2, it was revealed that L2 teachers, even though they might hold gender stereotypic beliefs, did not play a role in sustaining or legitimising any stereotypes that L2 learners might have (RQ 2). The results demonstrated that these beliefs were formed prior to university either by primary and secondary school teachers or families. Finally, in Study 3, since L2 male learners in the experimental conditions performed worse in the English listening test than the one in the control group, it was suggested that the phenomenon of stereotype threat might occur in the context of L2 learning as well. However, the effects of stereotype threat did not seem to be transferred through L2 reception self-efficacy and L2 listening anxiety (RQ3).

This programme of research demonstrates an original and rigorous approach to the issue of gender stereotyping of L2 learning in many ways. In addition to adopting an innovative mixed methods research design which increased the validity and reliability of the research, the study contributes to the literature by focusing on a context (Turkish EFL learners with a special focus on males) which has not gained much attention from researchers. As such, the results gained through the three studies gives researchers and language teachers important insights into the issue of gender stereotyping of L2 learning and its link to L2 performance via L2 anxiety and L2 self-efficacy.

References

- Aida, Y. (1994). Examination of horwitz, horwitz, and cope's construct of foreign language anxiety: The case of students of japanese. *The Modern Language Journal*, 78(2), 155-168. doi:10.1111/j.1540-4781.1994.tb02026.x
- Appel, M., Kronberger, N., & Aronson, J. (2011). Stereotype threat impairs ability building: Effects on test preparation among women in science and technology. *European Journal of Social Psychology*, 41(7), 904-913. doi:10.1002/ejsp.835
- Arbuckle, J. L. (1996). Full information estimation in the presence of incomplete data. In G. A. Marcoulides, & R. E. Schumacker (Eds.), *Advanced structural equation modeling: Issues and techniques* (pp. 277) Mahwah, NJ: Erlbaum.
- Archer, M., Bhaskar, R., Collier, A., Lawson, T., & Norrie, A. (Eds.). (2007). *Critical realism: Essential readings* Routledge.
- Armenta, B. E. (2010). Stereotype boost and stereotype threat effects: The moderating role of ethnic identification. *Cultural Diversity and Ethnic Minority Psychology*, 16(1), 94. doi:10.1037/a0017564
- Arnold, J., & Brown, H. D. (Eds.). (1999). *Introduction: A map of the terrain* Cambridge University Press.
- Aronson, J., Fried, C. B., & Good, C. (2002). Reducing the effects of stereotype threat on african american college students by shaping theories of intelligence. *Journal of Experimental Social Psychology*, 38(2), 113-125. doi:10.1006/jesp.2001.1491
- Aronson, J., Lustina, M. J., Good, C., Keough, K., Steele, C. M., & Brown, J. (1999). When white men can't do math: Necessary and sufficient factors in stereotype threat. *Journal of Experimental Social Psychology*, 35(1), 29-46. doi:10.1006/jesp.1998.1371
- Asakereh, A., & Dehghannezhad, M. (2015). Student satisfaction with EFL speaking classes: Relating speaking self-efficacy and skills achievement. *Issues in Educational Research*, 25(4), 345-363.
- Babbie, E. R. (2013). *The practice of social research* (13th ed.). CA: Wadsworth Cengage Learning.
- Bandura, A. (1986). *Social foundations of thought and action*. NJ: Prentice Hall: Englewood Cliffs.
- Bandura, A. (1997). *Self-efficacy: The exercise of control* New York: Freeman.
- Bandura, A. (2006). Guide for constructing self-efficacy scales. *Self-Efficacy Beliefs of Adolescents*, 5(307-337)

- Bandura, A., & Walters, R. H. (1963). *Social learning and personality development*. Holt Rinehart and Winston: New York.
- Barber, S. J., & Mather, M. (2013). Stereotype threat can reduce older adults' memory errors. *The Quarterly Journal of Experimental Psychology*, 66(10), 1888-1895. doi:10.1080/17470218.2013.840656
- Barone, C. (2011). Some things never change: Gender segregation in higher education across eight nations and three decades. *Sociology of Education*, 84(2), 157-176. doi:10.1177/0038040711402099
- Beasley, M. A., & Fischer, M. J. (2012). Why they leave: The impact of stereotype threat on the attrition of women and minorities from science, math and engineering majors. *Social Psychology of Education*, 15(4), 427-448. doi:10.1007/s11218-012-9185-3
- Bem, S. L. (1983). Gender schema theory and its implications for child development: Raising gender-aschematic children in a gender-schematic society. *Signs: Journal of Women in Culture and Society*, 8(4), 598-616. doi:10.1086/493998
- Bem, S. L. (1993). *The lenses of gender: Transforming the debate on sexual inequality* Yale University Press.
- Bergman, M. M. (2010). On concepts and paradigms in mixed methods research. *Journal of Mixed Methods Research*, 4(3), 171-175. doi:10.1177/1558689810376950
- Beyer, S. (2014). Why are women underrepresented in computer science? gender differences in stereotypes, self-efficacy, values, and interests and predictors of future CS course-taking and grades. *Computer Science Education*, 24(2-3), 153-192. doi:10.1080/08993408.2014.963363
- Bhaskar, R. (2007). Theorising ontology. In C. Lawson, J. S. Latsis & M. Nuno (Eds.), *Contributions to social ontology* (pp. 192) Routledge.
- Biesta, G. (2010). Pragmatism and the philosophical foundations of mixed methods research. In A. Tashakkori, & C. Teddlie (Eds.), *Sage handbook of mixed methods in social and behavioral research* (pp. 95-118) Sage Thousand Oaks, CA.
- Blaikie, N. (2010). *Designing social research* Polity.
- Blaikie, N., & Priest, J. (2017). *Social research: Paradigms in action* John Wiley & Sons.
- Blaxter, L., Hughes, C., & Tight, M. (2010). *How to research* McGraw-Hill Education (UK).

- Boratav, H. B., Fişek, G. O., & Ziya, H. E. (2014). Unpacking masculinities in the context of social change: Internal complexities of the identities of married men in turkey. *Men and Masculinities*, 17(3), 299-324. doi:10.1177/1097184X14539511
- Bosson, J. K., Haymovitz, E. L., & Pinel, E. C. (2004). When saying and doing diverge: The effects of stereotype threat on self-reported versus non-verbal anxiety. *Journal of Experimental Social Psychology*, 40(2), 247-255. doi:10.1016/S0022-1031(03)00099-4
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101. doi:10.1191/1478088706qp063oa
- Brinkmann, S. (2014). Unstructured and semi-structured interviewing . In P. Leavy (Ed.), *The oxford handbook of qualitative research* (pp. 277-299) Oxford Library of Psychology.
- Brinkmann, S., & Kvale, S. (2015). *Interviews: Learning the craft of qualitative research interviewing* (3rd ed.). Los Angeles: Sage Publications.
- Brislin, R. W. (1970). Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology*, 1(3), 185-216. doi:10.1177/135910457000100301
- Brown, R. P., & Day, E. A. (2006). The difference isn't black and white: Stereotype threat and the race gap on raven's advanced progressive matrices. *Journal of Applied Psychology*, 91(4), 979. doi:10.1037/0021-9010.91.4.979
- Brown, S. (2016). L2 listening. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 201) Taylor & Francis.
- Bryman, A. (2006). Integrating quantitative and qualitative research: How is it done? *Qualitative Research*, 6(1), 97-113. doi:10.1177/1468794106058877
- Bryman, A. (2007). Barriers to integrating quantitative and qualitative research. *Journal of Mixed Methods Research*, 1(1), 8-22. doi:10.1177/2345678906290531
- Bryman, A. (2008). Of methods and methodology. *Qualitative Research in Organizations and Management: An International Journal*, 3(2), 159-168. doi:10.1108/17465640810900568
- Burns, A., & Siegel, J. (2017). *International perspectives on teaching the four skills in ELT: Listening, speaking, reading, writing* Springer.
- Bussey, K., & Bandura, A. (1999). Social cognitive theory of gender development and differentiation. *Psychological Review*, 106(4), 676. doi:10.1037/0033-295X.106.4.676

- Cameron, J. E., & Lalonde, R. N. (2001). Social identification and gender-related ideology in women and men. *British Journal of Social Psychology*, 40(1), 59-77. doi:10.1348/014466601164696
- Cameron, R. (2011). Mixed methods research: The five ps framework. *Electronic Journal of Business Research Methods*, 9(2)
- Caner, A., Güven, C., Okten, C., & Sakalli, S. O. (2016). Gender roles and the education gender gap in turkey. *Social Indicators Research*, 129(3), 1231-1254.
- Carr, J. (2002). Why boys into languages won't go: The problematic gender agenda in languages education. *Babel*, 37(2), 4.
- Carr, J., & Pauwels, A. (2006). *Boys and foreign language learning: Real boys don't do languages*. Basingstoke & New York: Palgrave Macmillan.
- Chang, A. C. (2008). Sources of listening anxiety in learning english as a foreign language. *Perceptual and Motor Skills*, 106(1), 21-34. doi:10.2466/pms.106.1.21-34
- Chastain, K. (1975). Affective and ability factors in second language acquisition. *Language Learning*, 25(1), 153-161. doi:10.1111/j.1467-1770.1975.tb00115.x
- Chen, H., & Boore, J. R. (2010). Translation and back-translation in qualitative nursing research: Methodological review. *Journal of Clinical Nursing*, 19(1-2), 234-239. doi:10.1111/j.1365-2702.2009.02896.x
- Cheng, L., Klinger, D., Fox, J., Doe, C., Jin, Y., & Wu, J. (2014). Motivation and test anxiety in test performance across three testing contexts: The CAEL, CET, and GEPT. *TESOL Quarterly*, 48(2), 300-330. doi:10.1002/tesq.105
- Cheng, Y., Horwitz, E. K., & Schallert, D. L. (1999). Language anxiety: Differentiating writing and speaking components. *Language Learning*, 49(3), 417-446. doi:10.1111/0023-8333.00095
- Cheng, Y. (2004a). EFL students' writing anxiety: Sources and implications. *English Teaching & Learning*, 29(2), 41-62. doi:10.1016/j.jslw.2004.07.001
- Cheng, Y. (2004b). A measure of second language writing anxiety: Scale development and preliminary validation. *Journal of Second Language Writing*, 13(4), 313-335. doi:10.1016/j.jslw.2004.07.001
- Cheng, Y. (2017). Development and preliminary validation of four brief measures of L2 language-skill-specific anxiety. *System*, doi:10.1016/j.system.2017.06.009
- Cheryan, S., Siy, J. O., Vichayapai, M., Drury, B. J., & Kim, S. (2011). Do female and male role models who embody STEM stereotypes hinder women's anticipated

success in STEM? *Social Psychological and Personality Science*, 2(6), 656-664.
doi:10.1177/1948550611405218

Chodorow, N. (1978). *The reproduction of mothering: Psychoanalysis and the sociology of gender*. Berkeley, CA: University of California Press.

Christ, T. W. (2010). Teaching mixed methods and action research: Pedagogical, practical, and evaluative considerations. In A. Tashakkori, & C. Teddlie (Eds.), *Sage handbook of mixed methods in social & behavioral research* (pp. 643-676) Sage.

Chung, B. G., Ehrhart, M. G., Holcombe Ehrhart, K., Hattrup, K., & Solamon, J. (2010). Stereotype threat, state anxiety, and specific self-efficacy as predictors of promotion exam performance. *Group & Organization Management*, 35(1), 77-107. doi:10.1177/1059601109354839

Cohen, G. L., Garcia, J., Apfel, N., & Master, A. (2006). Reducing the racial achievement gap: A social-psychological intervention. *Science*, 313(5791), 1307-1310. doi:10.1126/science.1128317

Cohen, G. L., & Sherman, D. K. (2005). Stereotype threat and the social and scientific contexts of the race achievement gap. *The American Psychologist*, 3(60), 270-271. doi:10.1037/0003-066X.60.3.270

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences hillsdale* (2nd ed.). New York: USA: Lawrence Earlbaum Associates.

Cohen, L., Manion, L., & Morrison, K. (2013). *Research methods in education* Routledge.

Cohen, L., Manion, L., & Morrison, K. (2018). *Research methods in education* Routledge.

Collier, A. (1994). *Critical realism: An introduction to roy bhaskar's philosophy*. New York, NY: Verso.

Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis* Psychology Press.

Connelly, L. M. (2008). Pilot studies. *Medsurg Nursing*, 17(6), 411-413.

Coolican, H. (2014). *Research methods and statistics in psychology* Routledge.

Cornwell, S., & McKay, T. (1999). Measuring writing apprehension in japan. *Paper Presented at the Annual Meeting of the Japan Association of Language Teachers*, Retrieved from <http://files.eric.ed.gov/fulltext/ED433712.pdf>

Council of Europe. (2001). *Common european framework of reference for languages: Learning, teaching, assessment* Cambridge University Press.

- Council of Higher Education. (2018). Number of graduates by classification of fields of education and training. Retrieved from <https://istatistik.yok.gov.tr/>
- Creswell, J. W. (2014). *A concise introduction to mixed methods research* Sage Publications.
- Creswell, J. W., & Clark, V. L. P. (2017). *Designing and conducting mixed methods research* SAGE Publications.
- Creswell, J. W., & Creswell, D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* Sage.
- Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches* Sage publications.
- Crocker, J., Garcia, J. A., & Nuer, N. (2008). From egosystem to ecosystem in intergroup interactions: Implications for intergroup reconciliation. In Nadler, A., Malloy, T., & Fisher, J. D. (Eds.). *Social psychology of intergroup reconciliation: From violent conflict to peaceful co-existence*. (pp. 171-194) Oxford University Press.
- Croizet, J., & Claire, T. (1998). Extending the concept of stereotype threat to social class: The intellectual underperformance of students from low socioeconomic backgrounds. *Personality and Social Psychology Bulletin*, 24(6), 588-594. doi:10.1177/0146167298246003
- Crotty, M. (1998). *The foundations of social research: Meaning and perspective in the research process* Sage.
- Cumming, A. (2015). Theoretical orientations to L2 writing. In R. M. Manchón, & P. K. Matsuda (Eds.), *Handbook of second and foreign language writing* (pp. 65-90) Walter de Gruyter Boston, MA.
- Cvencek, D., Nasir, N. S., O'Connor, K., Wischnia, S., & Meltzoff, A. N. (2015). The development of math–race stereotypes: “They say chinese people are the best at math”. *Journal of Research on Adolescence*, 25(4), 630-637. doi:10.1111/jora.12151
- Daly, J. A., & Miller, M. D. (1975). The empirical development of an instrument to measure writing apprehension. *Research in the Teaching of English*, 9(3), 242-249.
- Danermark, B., Ekstrom, M., & Jakobsen, L. (2005). *Explaining society: An introduction to critical realism in the social sciences* Routledge.
- Davis, C., Aronson, J., & Salinas, M. (2006). Shades of threat: Racial identity as a moderator of stereotype threat. *Journal of Black Psychology*, 32(4), 399-417. doi:10.1177/0095798406292464

- Dawes, J. (2008). Do data characteristics change according to the number of scale points used. *International Journal of Market Research*, 50(1), 61-77.
- De Boer, H., Bosker, R. J., & van der Werf, Margaretha PC. (2010). Sustainability of teacher expectation bias effects on long-term student performance. *Journal of Educational Psychology*, 102(1), 168. doi:10.1037/a0017289
- De Raad, B. (2000). Differential psychology. *Encyclopedia of Psychology*, Oxford: American Psychological Association and Oxford University Press, 3, 41-44.
- Deemer, E. D., Thoman, D. B., Chase, J. P., & Smith, J. L. (2014). Feeling the threat stereotype threat as a contextual barrier to women's science career choice intentions. *Journal of Career Development*, 41(2), 141-158. doi:10.1177/0894845313483003
- Denzin, N. K. (2008). The new paradigm dialogs and qualitative inquiry. *International Journal of Qualitative Studies in Education*, 21(4), 315-325. doi:10.1080/09518390802136995
- Devine, A., Fawcett, K., Szűcs, D., & Dowker, A. (2012). Gender differences in mathematics anxiety and the relation to mathematics performance while controlling for test anxiety. *Behavioral and Brain Functions*, 8(1), 33. doi:10.1186/1744-9081-8-33
- Dewaele, J. (2013). The link between foreign language classroom anxiety and psychoticism, extraversion, and neuroticism among adult bi-and multilinguals. *The Modern Language Journal*, 97(3), 670-684. doi:10.1111/j.1540-4781.2013.12036.x
- Dewaele, J. (2017). Psychological dimensions and foreign language anxiety. In S. Loewen, & M. Sato (Eds.), *The routledge handbook of instructed second language acquisition* (pp. 433-450) Taylor & Francis Inc. doi:10.4324/9781315676968
- Dewaele, J., & MacIntyre, P. D. (2014). The two faces of janus? anxiety and enjoyment in the foreign language classroom. *Studies in Second Language Learning and Teaching*, 4(2), 237-274. doi:10.14746/ssllt.2014.4.2.5
- Dörnyei, Z. (2000). Motivation in action: Towards a process-oriented conceptualisation of student motivation. *British Journal of Educational Psychology*, 70(4), 519-538. doi:10.1348/000709900158281
- Dörnyei, Z. (2005). *The Psychology of the Language Learner: Individual Differences in Second Language Acquisition*. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.
- Dörnyei, Z., & Ryan, S. (2015). *The psychology of the language learner revisited*. New York, NY: Routledge. doi:10.4324/9781315779553

- Dörnyei, Z., & Taguchi, T. (2009). *Questionnaires in second language research: Construction, administration, and processing* Routledge.
- Duckworth, A. L., & Yeager, D. S. (2015). Measurement matters: Assessing personal qualities other than cognitive ability for educational purposes. *Educational Researcher*, 44(4), 237-251. doi: 10.3102/0013189X15584327
- Earley, M. A. (2007). Developing a syllabus for a mixed-methods research course. *International Journal of Social Research Methodology*, 10(2), 145-162. doi:10.1080/13645570701334118
- Eckert, P., & McConnell-Ginet, S. (1992). Think practically and look locally: Language and gender as community-based practice. *Annual Review of Anthropology*, 21, 461-490.
- Eisinga, R., Te Grotenhuis, M., & Pelzer, B. (2013). The reliability of a two-item scale: Pearson, cronbach, or spearman-brown? *International Journal of Public Health*, 58(4), 637-642. doi:10.1007/s00038-012-0416-3
- Elaldi, S. (2016). Foreign language anxiety of students studying english language and literature: A sample from turkey. *Educational Research and Reviews*, 11(6), 219. doi:10.5897/ERR2015.2507
- Elkhafaifi, H. (2005). Listening comprehension and anxiety in the arabic language classroom. *The Modern Language Journal*, 89(2), 206-220. doi:10.1111/j.1540-4781.2005.00275.x
- Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling*, 8(3), 430-457. doi:10.1207/S15328007SEM0803_5
- Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance: Attentional control theory. *Emotion*, 7(2), 336.
- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, 4(3), 272.
- Fagot, B. I., Rodgers, C. S., & Leinbach, M. D. (2000). Theories of gender socialization. In T. Eckes, & H. M. Trautner (Eds.), *The developmental social psychology of gender* (pp. 79-104) Psychology Press.
- Finnigan, K. M., & Corker, K. S. (2016). Do performance avoidance goals moderate the effect of different types of stereotype threat on women's math performance? *Journal of Research in Personality*, 63, 36-43. doi:10.1016/j.jrp.2016.05.009
- Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2012). *How to design and evaluate research in education*. New York: McGraw-Hill.

- Freshwater, D., & Cahill, J. (2013). Paradigms lost and paradigms regained. *Journal of Mixed Methods Research*, 7(1), 3-5. doi:10.1177/1558689812471276
- Fugard, A. J., & Potts, H. W. (2015). Supporting thinking on sample sizes for thematic analyses: A quantitative tool. *International Journal of Social Research Methodology*, 18(6), 669-684. doi:10.1080/13645579.2015.1005453
- Gan, Z. (2012). Understanding L2 speaking problems: Implications for ESL curriculum development in a teacher training institution in hong kong. *Australian Journal of Teacher Education*, 37(1), 3. doi:10.14221/ajte.2012v37n1.4
- Ganley, C. M., Mingle, L. A., Ryan, A. M., Ryan, K., Vasilyeva, M., & Perry, M. (2013). An examination of stereotype threat effects on girls' mathematics performance. *Developmental Psychology*, 49(10), 1886. doi:10.1037/a0031412
- Gardner, R. C., Day, J. B., & MacIntyre, P. D. (1992). Integrative motivation, induced anxiety, and language learning in a controlled environment. *Studies in Second Language Acquisition*, 14(2), 197-214. doi:10.1080/10615809708249294
- Gass, S. M., & Selinker, L. (2008). *Second language acquisition: An introductory course* Routledge.
- Gay, L. R., Mills, G. E., & Airasian, P. W. (2003). *Educational research: Competencies for analysis and applications, student value edition* Upper Saddle River, NJ: Merrill.
- Gilbert, D. T., & Hixon, J. G. (1991). The trouble of thinking: Activation and application of stereotypic beliefs. *Journal of Personality and Social Psychology*, 60(4), 509.
- Gill, P., Stewart, K., Treasure, E., & Chadwick, B. (2008). Methods of data collection in qualitative research: Interviews and focus groups. *British Dental Journal*, 204(6), 291. doi:10.1038/bdj.2008.192
- Gillham, B. (2007). *Developing a questionnaire* A&C Black.
- Gkonou, C. (2014). The sociolinguistic parameters of L2 speaking anxiety. *Classroom-oriented research* (pp. 15-32) Springer. doi:10.1007/978-3-319-00188-3_2
- Goh, C. C. (2000). A cognitive perspective on language learners' listening comprehension problems. *System*, 28(1), 55-75. doi:10.1016/S0346-251X(99)00060-3
- Goh, C. C. (2016). Cognition, metacognition, and L2 listening. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 214-228) Routledge New York, NY.

- Goh, C. C. (2017). Research into practice: Scaffolding learning processes to improve speaking performance. *Language Teaching*, 50(2), 247-260. doi:10.1017/S0261444816000483
- Goh, C. C., & Burns, A. (2012). *Teaching speaking: A holistic approach* Cambridge University Press.
- Gonzales, P. M., Blanton, H., & Williams, K. J. (2002). The effects of stereotype threat and double-minority status on the test performance of latino women. *Personality and Social Psychology Bulletin*, 28(5), 659-670. doi:10.1177/0146167202288010
- Gopang, I. B., Ansari, S., Kulsoom, U., & Laghari, A. (2017). An empirical investigation of foreign language anxiety in pakistani university. *International Journal of English Linguistics*, 7(2), 21. doi:10.5539/ijel.v7n2p21
- Gorenflo, D. W., & McConnell, J. V. (1991). The most frequently cited journal articles and authors in introductory psychology textbooks. *Teaching of Psychology*, 18(1), 8-12.
- Gorsuch, G. (2009). Investigating second language learner Self-Efficacy and future expectancy of second language use for High-Stakes program evaluation. *Foreign Language Annals*, 42(3), 505-540. doi:10.1111/j.1944-9720.2009.01034.x
- Grabe, W. (2010). Reading in a second language. In R. B. Kaplan (Ed.), *The oxford handbook of applied linguistics* (pp. 88-99) Oxford University Press.
- Grabe, W. (2011). L2 reading comprehension and development. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 299-311) Routledge.
- Graham, J. W., Van Horn, M. L., & Taylor, B. J. (2012). Dealing with the problem of having too many variables in the imputation model. In J. W. Graham (Ed.), *Missing data* (pp. 213-228). New York, NY: Springer.
- Graham, S. (2007). Learner strategies and self-efficacy: Making the connection. *Language Learning Journal*, 35(1), 81-93. doi:10.1080/09571730701315832
- Graham, S., & Macaro, E. (2008). Strategy instruction in listening for lower-intermediate learners of french. *Language Learning*, 58(4), 747-783. doi:10.1111/j.1467-9922.2008.00478.x
- Gray, D. E. (2014). *Doing research in the real world* (3rd ed.) Sage.
- Greene, J. C. (2007). *Mixed methods in social inquiry* John Wiley & Sons.
- Greene, J. C. (2008). Is mixed methods social inquiry a distinctive methodology? *Journal of Mixed Methods Research*, 2(1), 7-22. doi:10.1177/1558689807309969

- Greene, J. C., Caracelli, V. J., & Graham, W. F. (1989). Toward a conceptual framework for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis*, 11(3), 255-274. doi:10.3102/01623737011003255
- Greene, J. C., & Hall, J. (2010). Dialectics and pragmatism: Being of consequence. In A. Tashakkori, & C. Teddlie (Eds.), *SAGE handbook of mixed methods in social and behavioral research*, (2nd ed., pp. 119-143) Thousand Oaks, Sage.
- Gregersen, T., & Horwitz, E. K. (2002). Language learning and perfectionism: Anxious and non-anxious language learners' reactions to their own oral performance. *The Modern Language Journal*, 86(4), 562-570. doi:10.1111/1540-4781.00161
- Gregersen, T., MacIntyre, P. D., & Meza, M. D. (2014). The motion of emotion: Idiodynamic case studies of learners' foreign language anxiety. *The Modern Language Journal*, 98(2), 574-588. doi:10.1111/modl.12084
- Guba, E. G., & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. In N. K. Denzin, & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 105) Sage, Thousand Oaks, CA.
- Gunderson, E. A., Ramirez, G., Levine, S. C., & Beilock, S. L. (2012). The role of parents and teachers in the development of gender-related math attitudes. *Sex Roles*, 66(3-4), 153-166. doi:10.1007/s11199-011-9996-2
- Hagtvet, K. A., & Benson, J. (1997). The motive to avoid failure and test anxiety responses: Empirical support for integration of two research traditions. *Anxiety, Stress, and Coping*, 10(1), 35-57. doi:10.1080/10615809708249294
- Hall, R. (2013). Mixed methods: In search of a paradigm. In T. Le, & Q. Le (Eds.), *Conducting research in a changing and challenging world* (pp. 71-78) Laboratory and Clinical Research.
- Hamzaoğlu, H., & Koçoğlu, Z. (2016). The application of podcasting as an instructional tool to improve turkish EFL learners' speaking anxiety. *Educational Media International*, 53(4), 313-326. doi:10.1080/09523987.2016.1254889
- Han, J., & Hiver, P. (2018). Genre-based L2 writing instruction and writing-specific psychological factors: The dynamics of change. *Journal of Second Language Writing*, 40, 44-59. doi:10.1016/j.jslw.2018.03.001
- Haslam, S. A., & McGarty, C. (2014). *Research methods and statistics in psychology* Sage.
- Hayes, A. F. (2013). Introduction to mediation, moderation, and conditional process analysis: A regression-based approach. new york, NY: The guilford press.
- Hayes, A. F. (2017). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach* Guilford Publications.

- Hayes, A. F., & Preacher, K. J. (2014). Statistical mediation analysis with a multicategorical independent variable. *British Journal of Mathematical and Statistical Psychology*, 67(3), 451-470. doi:10.1111/bmsp.12028
- Hesse-Biber, S. N., & Johnson, R. B. (2015). *The oxford handbook of multimethod and mixed methods research inquiry* Oxford University Press.
- Hill, C., Corbett, C., & St Rose, A. (2010). *Why so few? women in science, technology, engineering, and mathematics*. Washington, DC: AAUW.
- Hill, F., Mammarella, I. C., Devine, A., Caviola, S., Passolunghi, M. C., & Szűcs, D. (2015). Maths anxiety in primary and secondary school students: Gender differences, developmental changes and anxiety specificity. *Learning and Individual Differences*, 48, 45-53. doi:10.1016/j.lindif.2016.02.006
- Hogg, M. A. (2006). Social identity theory. In P. J. Burke (Ed.), *Contemporary social psychological theories* (pp. 111-1369). Stanford, CA: Stanford University Press.
- Hole, G. (2012). Experimental design. In G. M. Breakwell, J. A. Smith & D. B. Wright (Eds.), *Research methods in psychology* (pp. 39-73). London: Sage.
- Horwitz, E. K. (1986). Preliminary evidence for the reliability and validity of a foreign language anxiety scale. *Tesol Quarterly*, 20(3), 559-562. doi:10.2307/3586302
- Horwitz, E. K. (2000). It ain't over'til it's over: On foreign language anxiety, first language deficits, and the confounding of variables. *The Modern Language Journal*, 84(2), 256-259. doi:10.1111/0026-7902.00067
- Horwitz, E. K. (2001). Language anxiety and achievement. *Annual Review of Applied Linguistics*, 21, 112-126. doi:10.1017/S0267190501000071
- Horwitz, E. K. (2010). Foreign and second language anxiety. *Language Teaching*, 43(02), 154-167. doi:10.1017/S026144480999036X
- Horwitz, E. K. (2017). On the misreading of horwitz, horwitz, and cope (1986) and the need to balance anxiety research and the experiences of anxious language learners. In C. Gkonou, M. Daubney & J. Dewaele (Eds.), *New insights into language anxiety: Theory, research and educational implications*. bristol, england: Multilingual matters (pp. 31-47)
- Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *The Modern Language Journal*, 70(2), 125-132. doi:10.1111/j.1540-4781.1986.tb05256.x
- Horwitz, E. K., Tallon, M., & Luo, H. (2010). Foreign language anxiety. In J. C. Cassady (Ed.), *Anxiety in schools: The causes, consequences, and solutions for academic anxieties* (pp. 95-115). New York: Peter Lang.

- Horwitz, E. K., & Young, D. J. (1991). *Language anxiety: From theory and research to classroom implications*. Engelwood Cliffs, NJ: Prentice-Hall.
- Hoyt, C. L., & Murphy, S. E. (2016). Managing to clear the air: Stereotype threat, women, and leadership. *The Leadership Quarterly*, doi:10.1016/j.leaqua.2015.11.002
- Hsieh, P. P., & Kang, H. (2010). Attribution and self-efficacy and their interrelationship in the Korean EFL context. *Language Learning*, 60(3), 606-627. doi:10.1111/j.1467-9922.2010.00570.x
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1-55. doi:10.1080/10705519909540118
- Huston, A. C. (1983). Sex-typing. *Handbook of Child Psychology*, 4, 387-467.
- In'nami, Y. (2006). The effects of test anxiety on listening test performance. *System*, 34(3), 317-340. doi:10.1016/j.system.2006.04.005
- Inzlicht, M., & Schmader, T. (2012). *Stereotype threat: Theory, process, and application* Oxford University Press.
- Jamieson, J. P., & Harkins, S. G. (2007). Mere effort and stereotype threat performance effects. *Journal of Personality and Social Psychology*, 93(4), 544. doi:10.1037/0022-3514.93.4.544
- Jamieson, J. P., & Harkins, S. G. (2012). Distinguishing between the effects of stereotype priming and stereotype threat on math performance. *Group Processes & Intergroup Relations*, 15(3), 291-304. doi:10.1177/1368430211417833
- Jamshidnejad, A. (2011). Functional approach to communication strategies: An analysis of language learners' performance in interactional discourse. *Journal of Pragmatics*, 43(15), 3757-3769. doi:10.1016/j.pragma.2011.09.017
- Jeon, E. H., & Yamashita, J. (2014). L2 reading comprehension and its correlates: A meta-analysis. *Language Learning*, 64(1), 160-212. doi:10.1111/lang.12034
- Jin, Y., de Bot, K., & Keijzer, M. (2017). Affective and situational correlates of foreign language proficiency: A study of Chinese university learners of English and Japanese. *Studies in Second Language Learning and Teaching*, 7(1), 105-125. doi:10.14746/ssllt.2017.7.1.6
- Jinks, J., & Morgan, V. (1999). Children's perceived academic self-efficacy: An inventory scale. *The Clearing House*, 72(4), 224-230. doi:10.1080/00098659909599398

- Johns, M., Schmader, T., & Martens, A. (2005). Knowing is half the battle teaching stereotype threat as a means of improving women's math performance. *Psychological Science*, 16(3), 175-179. doi:10.1111/j.0956-7976.2005.00799.x
- Johnson, B., & Christensen, L. (2000). *Educational research: Quantitative and qualitative approaches*. Allyn & Bacon.
- Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, 33(7), 14-26. doi:10.3102/0013189X033007014
- Johnson, R. B., Onwuegbuzie, A. J., & Turner, L. A. (2007). Toward a definition of mixed methods research. *Journal of Mixed Methods Research*, 1(2), 112-133. doi:10.1177/1558689806298224
- Joy, J. L. (2013). The altitude of test anxiety among second language learners. *Language Testing in Asia*, 3(1), 10. doi:10.1186/2229-0443-3-10
- Kao, P., Chen, K. T., & Craigie, P. (2017). Gender differences in strategies for coping with foreign language learning anxiety. *Social Behavior and Personality: An International Journal*, 45(2), 205-210. doi:10.2224/sbp.5771
- Katz, I. (1964). Review of evidence relating to effects of desegregation on the intellectual performance of negroes. *American Psychologist*, 19(6), 381. doi:10.1037/h0040569
- Katz, I., Epps, E. G., & Axelson, L. J. (1964). Effect upon negro digit-symbol performance of anticipated comparison with whites and with other negroes. *The Journal of Abnormal and Social Psychology*, 69(1), 77.
- Katz, I., Roberts, S. O., & Robinson, J. M. (1965). Effects of task difficulty, race of administrator, and instructions on digit-symbol performance of negroes. *Journal of Personality and Social Psychology*, 2(1), 53. doi:10.1037/h0022080
- Kaye, L. K., & Pennington, C. R. (2016). "Girls can't play": The effects of stereotype threat on females' gaming performance. *Computers in Human Behavior*, 59, 202-209. doi:10.1016/j.chb.2016.02.020
- Kaye, L. K., Pennington, C. R., & McCann, J. J. (2018). Do casual gaming environments evoke stereotype threat? examining the effects of explicit priming and avatar gender. *Computers in Human Behavior*, 78, 142-150. doi:10.1016/j.chb.2017.09.031
- Keller, J. (2002). Blatant stereotype threat and women's math performance: Self-handicapping as a strategic means to cope with obtrusive negative performance expectations. *Sex Roles*, 47(3-4), 193-198. doi:10.1003307511

- Keller, J. (2007). Stereotype threat in classroom settings: The interactive effect of domain identification, task difficulty and stereotype threat on female students' maths performance. *British Journal of Educational Psychology*, 77(2), 323-338. doi:10.1348/000709906X113662
- Keller, J., & Dauenheimer, D. (2003). Stereotype threat in the classroom: Dejection mediates the disrupting threat effect on women's math performance. *Personality and Social Psychology Bulletin*, 29(3), 371-381. doi:10.1177/0146167202250218
- Kendall, L. (2008). The conduct of qualitative interviews: Research questions, methodological issues, and researching online. *Handbook of research on new literacies* (pp. 133-149) Routledge.
- Kiefer, A. K., & Sekaquaptewa, D. (2007). Implicit stereotypes and women's math performance: How implicit gender-math stereotypes influence women's susceptibility to stereotype threat. *Journal of Experimental Social Psychology*, 43(5), 825-832. doi:10.1016/j.jesp.2006.08.004
- Kim, J. (2000). *Foreign language listening anxiety: A study of korean students learning english*. University of Texas, Austin: Unpublished doctoral dissertation.
- Kim, J. (2005). The reliability and validity of a foreign language listening anxiety scale. *Korean Journal of English Language and Linguistics*, 5(2), 213-235.
- Kim, Y. (2002). Construction of a theoretical model of foreign language anxiety and development of a measure of the construct: Korean university EFL learners' case. *Unpublished Doctoral Dissertation, Indiana University*,
- Kleinmann, H. H. (1977). Avoidance behavior in adult second language acquisition. *Language Learning*, 27(1), 93-107. doi:10.1111/j.1467-1770.1977.tb00294.x
- Koenig, A. M., & Eagly, A. H. (2005). Stereotype threat in men on a test of social sensitivity. *Sex Roles*, 52(7-8), 489-496. doi:10.1007/s11199-005-3714-x
- Kohlberg, L. (1966). A cognitive-developmental analysis of children's sex-role concepts and attitudes. In E. E. Maccoby (Ed.), (pp. 82-123). Stanford, CA: Stanford University Press.
- Krashen, S. D. (1982). *Principles and practice in second language acquisition*. Oxford: Pergamon.
- Kray, L. J., Thompson, L., & Galinsky, A. (2001). Battle of the sexes: Gender stereotype confirmation and reactance in negotiations. *Journal of Personality and Social Psychology*, 80(6), 942. doi:10.1037//0022-3514.80.6.94 2
- Krosnick, J. A. (2018). Questionnaire design. *The palgrave handbook of survey research* (pp. 439-455) Springer.

- Kuhn, T. S. (1970). *The structure of scientific revolutions*. Chicago: University of Chicago Press.
- Kunt, N. (1999). Anxiety and beliefs about language learning: A study of turkish-speaking university students learning english in north cyprus.
- Kvam, K. E., Schearing, L. S., & Ziegler, V. H. (1999). *Eve and adam: Jewish, christian, and muslim readings on genesis and gender* Indiana University Press.
- Levy, B. (1996). Improving memory in old age through implicit self-stereotyping. *Journal of Personality and Social Psychology*, 71(6), 1092. doi:10.1037/0022-3514.71.6.1092
- Li, W., & Renandya, W. A. (2012). Effective approaches to teaching listening: Chinese EFL teachers' perspectives. *The Journal of Asia TEFL*, 9(4), 79-111.
- Li, Y., & Wang, C. (2010). An empirical study of reading self-efficacy and the use of reading strategies in the chinese EFL context. *Asian EFL Journal*, 12(2), 144-162.
- Liebert, R. M., & Morris, L. W. (1967). Cognitive and emotional components of test anxiety: A distinction and some initial data. *Psychological Reports*, 20(3), 975-978. doi:10.2466/pr0.1967.20.3.975
- Lightbown, P. M., & Spada, N. (2013). *How languages are learned* (4th Edition). Oxford: Oxford University Press.
- Liu, M. (2013). English bar as a venue to boost students' speaking self-efficacy at the tertiary level. *English Language Teaching*, 6(12), 27-37. doi:10.5539/elt.v6n12p27
- Loewen, S. (2014). *Introduction to instructed second language acquisition* Routledge.
- Lu, H., & Luk, J. (2014). "I would study harder if I was a girl": Gendered narratives of low-achieving male and high-achieving female EFL learners. *Journal of Language, Identity & Education*, 13(1), 1-15. doi:10.1080/15348458.2013.835571
- Lu, Z., & Liu, M. (2011). Foreign language anxiety and strategy use: A study with chinese undergraduate EFL learners. *Journal of Language Teaching & Research*, 2(6) doi:10.4304/jltr.2.6.1298-1305
- Luo, H. (2012). Sources of foreign language anxiety: Towards a four-dimensional model. *Contemporary Foreign Language Studies*, 12, 49-61.
- MacIntyre, P. D. (2012). The idiodynamic method: A closer look at the dynamics of communication traits. *Communication Research Reports*, 29(4), 361-367. doi:10.1080/08824096.2012.723274
- MacIntyre, P. D. (2017). An overview of language anxiety research and trends in its development. In C. Gkonou, M. Daubney & J. Dewaele (Eds.), *New insights into*

language anxiety: Theory, research and educational implications.bristol: Multilingual matters (pp. 11-30)

MacIntyre, P. D., & Gardner, R. C. (1989). Anxiety and second-language learning: Toward a theoretical clarification. *Language Learning*, 39(2), 251-275. doi:10.1111/j.1467-1770.1989.tb00423.x

MacIntyre, P. D., & Serroul, A. (2015). Motivation on a per-second timescale: Examining approach-avoidance motivation during L2 task performance. In Z. Dörnyei, A. Henry & P. D. MacIntyre (Eds.), *Motivational dynamics in language learning* (pp. 109-138). New York, NY: Multilingual Matters.

MacKinnon, D. (2012). *Introduction to statistical mediation analysis* Routledge.

Magogwe, J. M., & Oliver, R. (2007). The relationship between language learning strategies, proficiency, age and self-efficacy beliefs: A study of language learners in botswana. *System*, 35(3), 338-352. doi:10.1016/j.system.2007.01.003

Mak, B. (2011). An exploration of speaking-in-class anxiety with chinese ESL learners. *System*, 39(2), 202-214. doi:10.1016/j.system.2011.04.002

Manchón, R. (2009). *Writing in foreign language contexts: Learning, teaching, and research* Multilingual Matters.

Marsh, H. W., Pekrun, R., Parker, P. D., Murayama, K., Guo, J., Dicke, T., & Arens, A. K. (2018). The murky distinction between self-concept and self-efficacy: Beware of lurking jingle-jangle fallacies. *Journal of Educational Psychology*, doi:10.1037/edu0000281

Martin, C. L. (1990). Attitudes and expectations about children with nontraditional and traditional gender roles. *Sex Roles*, 22(3-4), 151-166.

Martin, S., & Valdivia, I. M. A. (2017). Students' feedback beliefs and anxiety in online foreign language oral tasks. *International Journal of Educational Technology in Higher Education*, 14(1), 18. doi:10.1186/s41239-017-0056-z

Marx, D. M., & Roman, J. S. (2002). Female role models: Protecting women's math test performance. *Personality and Social Psychology Bulletin*, 28(9), 1183-1193. doi:10.1177/01461672022812004

Marx, D. M., Stapel, D. A., & Muller, D. (2005). We can do it: The interplay of construal orientation and social comparisons under threat. *Journal of Personality and Social Psychology*, 88(3), 432. doi:10.1037/0022-3514.88.3.432

Maxwell, J. A. (2012). *A realist approach for qualitative research* Sage.

- Maxwell, J. A., & Mittapalli, K. (2010). Realism as a stance for mixed methods research. In A. Tashakkori, & C. Teddlie (Eds.), *Handbook of mixed methods in social & behavioral research* (pp. 145-168) Sage Thousand Oaks, CA.
- Mayer, D. M., & Hanges, P. J. (2003). Understanding the stereotype threat effect with "culture-free" tests: An examination of its mediators and measurement. *Human Performance*, 16(3), 207-230. doi:10.1207/S15327043HUP1603_3
- McAuliffe, M., & Brooks, G. (2018). The development of a listening course for Japanese university students. In A. Burns, & J. Siegel (Eds.), *International perspectives on teaching the four skills in ELT* (pp. 51-63) Springer.
- McDermott, M. A. N., & Palchanes, K. (1994). A literature review of the critical elements in translation theory. *Journal of Nursing Scholarship*, 26(2), 113-118. doi:10.1111/j.1547-5069.1994.tb00928.x
- Meijer, J. (2001). Learning potential and anxious tendency: Test anxiety as a bias factor in educational testing. *Anxiety, Stress and Coping*, 14(3), 337-362. doi:10.1080/10615800108248361
- Mercer, S., & Williams, M. (2014). *Multiple perspectives on the self in SLA* Multilingual Matters. doi:10.14746/sslit.2015.5.1.10
- Mertens, D. M. (2003). Mixed methods and the politics of human research: The transformative-emancipatory perspective. In A. Tashakkori, & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research* (pp. 135-164)
- Mertens, D. M. (2007). Transformative paradigm: Mixed methods and social justice. *Journal of Mixed Methods Research*, 1(3), 212-225. doi:10.1177/1558689807302811
- Mertens, D. M. (2010). Philosophy in mixed methods teaching: The transformative paradigm as illustration. *International Journal of Multiple Research Approaches*, 4(1), 9-18. doi:10.5172/mra.2010.4.1.009
- Mertens, D. M. (2014). *Research and evaluation in education and psychology: Integrating diversity with quantitative, qualitative, and mixed methods* Sage publications.
- Miller, D. I., Eagly, A. H., & Linn, M. C. (2015). Women's representation in science predicts national gender-science stereotypes: Evidence from 66 nations. *Journal of Educational Psychology*, 107(3), 631. doi:10.1037/edu0000005
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63(2), 81.

- Mills, N. (2014). Self-efficacy in second language acquisition. In S. Mercer, & M. Williams (Eds.), *Multiple perspectives on the self in SLA* (pp. 6-22) Multilingual Matters Bristol.
- Mills, N., Pajares, F., & Herron, C. (2006). A reevaluation of the role of anxiety: Self-efficacy, anxiety, and their relation to reading and listening proficiency. *Foreign Language Annals*, 39(2), 276-295. doi:10.1111/j.1944-9720.2006.tb02266.x
- Mills, N., & Peron, M. (2009). Global simulation and writing self-beliefs of intermediate french students. *International Journal of Applied Linguistics: Special Issue on Learning and Teaching L2 Writing*, 156, 239-273. doi:10.2143/ITL.156.0.2034436
- Mischel, W. (1970). Sex-typing and socialization. *Carmichael's Manual of Child Psychology*, 2, 3-72.
- Moeller, A. K., & Catalano, T. (2015). Foreign language teaching and learning. In J. D. Wright (Ed.), *International encyclopedia of the social & behavioral sciences* (2nd ed., pp. 327-332). Oxford: Elsevier.
- Mooney, A., & Evans, B. (2015). *Language, society and power: An introduction* Routledge.
- Morgan, D. L. (2007). Paradigms lost and pragmatism regained: Methodological implications of combining qualitative and quantitative methods. *Journal of Mixed Methods Research*, 1(1), 48-76. doi:10.1177/2345678906292462
- Morgan, D. L. (2014). Pragmatism as a paradigm for social research. *Qualitative Inquiry*, 20(8), 1045-1053. doi:10.1177/1077800413513733
- Murtiningsih, S. R., & Hapsari, W. (2018). Teaching reading to encourage critical thinking and collaborative work. In A. Burns, & J. Siegel (Eds.), *International perspectives on teaching the four skills in ELT* (pp. 153-165) Springer.
- Muthén, L. K., & Muthén, B. O. (2012). Mplus version 7 user's guide. Los Angeles, CA: Muthén & Muthén,
- Muthén, L. K., & Muthén, B. O. (2013). Mplus 7.11. Los Angeles, CA: Muthén & Muthén.
- Nadler, J. T., & Clark, M. H. (2011). Stereotype threat: A Meta-Analysis comparing african americans to hispanic americans 1. *Journal of Applied Social Psychology*, 41(4), 872-890. doi:10.1111/j.1559-1816.2011.00739.x
- Nakatani, Y. (2005). The effects of awareness-raising training on oral communication strategy use. *The Modern Language Journal*, 89(1), 76-91. doi:10.1111/j.0026-7902.2005.00266.x

- Nosratinia, M., Saveiy, M., & Zaker, A. (2014). EFL learners' self-efficacy, metacognitive awareness, and use of language learning strategies: How are they associated? *Theory and Practice in Language Studies*, 4(5), 1080. doi:10.4304/tpls.4.5.1080-1092
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychological theory*. New York, NY: MacGraw-Hill,
- O'Brien, L. T., & Crandall, C. S. (2003). Stereotype threat and arousal: Effects on women's math performance. *Personality and Social Psychology Bulletin*, 29(6), 782-789. doi:10.1177/0146167203029006010
- O'Cathain, A., Murphy, E., & Nicholl, J. (2008). The quality of mixed methods studies in health services research. *Journal of Health Services Research & Policy*, 13(2), 92-98. doi:10.1258/jhsrp.2007.007074
- Osborne, J. W. (2001). Testing stereotype threat: Does anxiety explain race and sex differences in achievement? *Contemporary Educational Psychology*, 26(3), 291-310. doi:10.1006/ceps.2000.1052
- Owens, J., & Massey, D. S. (2011). Stereotype threat and college academic performance: A latent variables approach. *Social Science Research*, 40(1), 150-166. doi:10.1016/j.ssresearch.2010.09.010
- Oxford, R. L. (1993). Research update on teaching L2 listening. *System*, 21(2), 205-211. doi:10.1016/0346-251X(93)90042-F
- Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research*, 66(4), 543-578.
- Pajares, F., Hartley, J., & Valiante, G. (2001). Response format in writing self-efficacy assessment: Greater discrimination increases prediction. *Measurement and Evaluation in Counseling and Development*, 33(4), 214.
- Pajares, F., & Schunk, D. (2001). The development of academic self-efficacy. *Development of Achievement Motivation. United States*, 7 doi:10.1016/B978-012750053-9/50003-6
- Paladino, M., Poddesu, L., Rauzi, M., Vaes, J., Cadinu, M., & Forer, D. (2009). Second language competence in the italian-speaking population of alto adige/sdtirol evidence for linguistic stereotype threat. *Journal of Language and Social Psychology*, 28(3), 222-243. doi:10.1177/0261927X09335333
- Panayides, P., & Walker, M. J. (2013). Evaluating the psychometric properties of the foreign language classroom anxiety scale for cypriot senior high school EFL students: The rasch measurement approach. *Europe's Journal of Psychology*, 9(3), 493-516. doi:10.5964/ejop.v9i3.611

- Park, G. (2014). Factor analysis of the foreign language classroom anxiety scale in Korean learners of English as a foreign language. *Psychological Reports*, 115(1), 261-275. doi:10.2466/28.11.PR0.115c10z2
- Park, H., & Lee, A. R. (2005). L2 learners' anxiety, self-confidence and oral performance. In *Proceedings of the 10th Conference of Pan-Pacific Association of Applied Linguistics* (pp. 107-208).
- Pavlenko, A. (2013). Language desire and 'Commodification of affect. In D. Gabrys-Barker, & J. Bielska (Eds.), *The affective dimension in second language acquisition* (pp. 3-28). Bristol: Multilingual Matters.
- Pearson, C. M. (2017). Real boys don't do language and literacy--or do they? *MITESOL Journal: An Online Publication of MITESOL*, 1(1), 3. Retrieved from <https://scholarworks.gvsu.edu/mitesol/vol1/iss1/3>
- Pekrun, R., Goetz, T., Frenzel, A. C., Barchfeld, P., & Perry, R. P. (2011). Measuring emotions in students' learning and performance: The achievement emotions questionnaire (AEQ). *Contemporary Educational Psychology*, 36(1), 36-48. doi:10.1016/j.cedpsych.2010.10.002
- Pekrun, R., Götz, T., & Perry, R. P. (2005). Achievement emotions questionnaire (AEQ). user's manual. *Department of Psychology, University of Munich, Munich, Germany*,
- Pennington, C. R. (2016). *The effects of stereotype threat on females' mathematical performance: A multi-faceted situational phenomenon? (doctoral dissertation)* Retrieved from <https://repository.edgehill.ac.uk/8295/>
- Pennington, C. R., & Heim, D. (2016). Creating a critical mass eliminates the effects of stereotype threat on women's mathematical performance. *British Journal of Educational Psychology*, 3(86), 353-368. doi:10.1111/bjep.12110
- Pennington, C. R., Heim, D., Levy, A. R., & Larkin, D. T. (2016). Twenty years of stereotype threat research: A review of psychological mediators. *PloS One*, 11(1), e0146487. doi:10.1371/journal.pone.0146487
- Pennington, C. R., Kaye, L. K., & McCann, J. J. (2018). Applying the multi-threat framework of stereotype threat in the context of digital gaming. *PloS One*, 13(2), e0192137. doi:10.1371/journal.pone.0192137
- Phillips, E. M. (1992). The effects of language anxiety on students' oral test performance and attitudes. *The Modern Language Journal*, 76(1), 14-26. doi:10.1111/j.1540-4781.1992.tb02573.x
- Picho, K., Rodriguez, A., & Finnie, L. (2013). Exploring the moderating role of context on the mathematics performance of females under stereotype threat: A meta-

- analysis. *The Journal of Social Psychology*, 153(3), 299-333.
doi:10.1080/00224545.2012.737380
- Piller, I., & Pavlenko, A. (2007). Globalization, gender, and multilingualism. In H. Decke-Cornill, & L. Volkmann (Eds.), *Gender studies and foreign language teaching* (pp. 15-30)
- Plante, I., Theoret, M., & Favreau, O. E. (2009). Student gender stereotypes: Contrasting the perceived maleness and femaleness of mathematics and language. *Educational Psychology*, 29(4), 385-405.
doi:10.1080/01443410902971500
- Polio, C. (2017). Second language writing development: A research agenda. *Language Teaching*, 50(2), 261-275. doi:10.1017/S0261444817000015
- Pomerantz, A. (2008). "Señora, boys just cannot speak spanish": Negotiating masculine identities in a foreign language classroom. *Working Papers in Educational Linguistics (WPEL)*, 23(2), 1.
- Preacher, K. J., Rucker, D. D., & Hayes, A. F. (2007). Addressing moderated mediation hypotheses: Theory, methods, and prescriptions. *Multivariate Behavioral Research*, 42(1), 185-227. doi:10.1080/00273170701341316
- Putwain, D. W. (2008). Deconstructing test anxiety. *Emotional and Behavioural Difficulties*, 13(2), 141-155. doi:10.1080/13632750802027713
- Renandya, W. A., & Hu, G. (2018). L2 listening in china: An examination of current practice. In A. Burns, & J. Siegel (Eds.), *International perspectives on teaching the four skills in ELT* (pp. 37-50) Springer.
- Retelsdorf, J., Schwartz, K., & Asbrock, F. (2015). "Michael can't read!" teachers' gender stereotypes and boys' reading self-concept. *Journal of Educational Psychology*, 107(1), 186. doi:10.1037/a0037107
- Riazi, A. (2007). Language learning strategy use: Perceptions of female arab english majors. *Foreign Language Annals*, 40(3), 433-440. doi:10.1111/j.1944-9720.2007.tb02868.x
- Riazi, M., Shi, L., & Haggerty, J. (2018). Analysis of the empirical research in the journal of second language writing at its 25th year (1992–2016). *Journal of Second Language Writing*, 41, 41-54. doi:0.1016/j.jslw.2018.07.002
- Robinson-Cimpian, J. P., Lubienski, S. T., Ganley, C. M., & Copur-Gencturk, Y. (2014). Teachers' perceptions of students' mathematics proficiency may exacerbate early gender gaps in achievement. *Developmental Psychology*, 50(4), 1262.
doi:10.1037/a0035073

- Ruane, J. M. (2005). *Essentials of research methods: A guide to social science research* Blackwell publishing.
- Ryan, K. E., & Ryan, A. M. (2005). Psychological processes underlying stereotype threat and standardized math test performance. *Educational Psychologist*, 40(1), 53-63. doi:10.1207/s15326985ep4001_4
- Ryle, R. (2011). *Questioning gender: A sociological exploration* Sage Publications.
- Saito, Y., Garza, T. J., & Horwitz, E. K. (1999). Foreign language reading anxiety. *The Modern Language Journal*, 83(2), 202-218. doi:10.1111/0026-7902.00016
- Saito, Y., & Samimy, K. K. (1996). Foreign language anxiety and language performance: A study of learner anxiety in beginning, intermediate, and advanced-level college students of Japanese. *Foreign Language Annals*, 29(2), 239-249. doi:10.1111/j.1944-9720.1996.tb02330.x
- Santos, A., Gorter, D., & Cenoz, J. (2017). Communicative anxiety in the second and third language. *International Journal of Multilingualism*, 14(1), 23-37. doi:10.1080/14790718.2017.1258980
- Santos, D., & Graham, S. (2018). What teachers say about listening and its pedagogy: A comparison between two countries. In A. Burns, & J. Siegel (Eds.), *International perspectives on teaching the four skills in ELT* (pp. 21-35) Springer.
- Sarason, I. G. (1961). Test anxiety and the intellectual performance of college students. *Journal of Educational Psychology*, 52(4), 201. doi:10.1037/h0049095
- Sarason, I. G. (Ed.). (1978). *The test anxiety scale: Concept and research*. Washington DC: Hemisphere/Wiley.
- Sarason, S. B., & Mandler, G. (1952). Some correlates of test anxiety. *The Journal of Abnormal and Social Psychology*, 47(4), 810. doi:10.1037/h0060009
- Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis.
- Schmader, T. (2002). Gender identification moderates stereotype threat effects on women's math performance. *Journal of Experimental Social Psychology*, 38(2), 194-201. doi:10.1006/jesp.2001.1500
- Schmader, T., & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality and Social Psychology*, 85(3), 440. doi:10.1037/0022-3514.85.3.440
- Schmader, T., Johns, M., & Barquissau, M. (2004). The costs of accepting gender differences: The role of stereotype endorsement in women's experience in the math domain. *Sex Roles*, 50(11-12), 835-850. doi:SERS.0000029101.74557.a0

- Schmenk, B. (2004). Language learning: A feminine domain? the role of stereotyping in constructing gendered learner identities. *TESOL Quarterly*, 38(3), 514-524. doi:10.2307/3588352
- Schmenk, B. (2007). Foreign language research and the feminization of language learning. In H. Decke-Cornill, & L. Volkmann (Eds.), *Gender studies and foreign language teaching* (pp. 121-135). Tübingen: Germany: Gunter Narr Verlag Tübingen.
- Schunk, D. H. (Ed.). (2012). *Social cognitive theory*. Washington, DC, US: American Psychological Association. doi:10.1037/13273-000
- Schunk, D. H., & DiBenedetto, M. K. (2016). Self-efficacy theory in education. In K. R. Wentzel, & D. B. Miele (Eds.), *Handbook of motivation at school* (pp. 34-54). New York, NY: Routledge.
- Schunk, D. H., & Pajares, F. (2009). Self-efficacy theory. In K. R. Wentzel, & A. Wigfield (Eds.), *Handbook of motivation at school* (pp. 35-53). New York, NY: Routledge.
- Schutz, P. A., & Pekrun, R. E. (2007). *Emotion in education*. Elsevier Academic Press.
- Scott, S. V., & Rodriguez, L. F. (2015). "A fly in the ointment" african american male preservice teachers' experiences with stereotype threat in teacher education. *Urban Education*, 50(6), 689-717. doi:10.1177/0042085913519335
- Scovel, T. (1978). The effect of affect on foreign language learning: A review of the anxiety research. *Language Learning*, 28(1), 129-142. doi:10.1111/j.1467-1770.1978.tb00309.x
- Seidman, I. (2013). *Interviewing as qualitative research: A guide for researchers in education and the social sciences* Teachers college press.
- Sellers, V. D. (2000). Anxiety and reading comprehension in spanish as a foreign language. *Foreign Language Annals*, 33(5), 512-520. doi:10.1111/j.1944-9720.2000.tb01995.x
- Shannon-Baker, P. (2016). Making paradigms meaningful in mixed methods research. *Journal of Mixed Methods Research*, 10(4), 319-334. doi:10.1177/1558689815575861
- Shapiro, J. R. (2011). Different groups, different threats: A multi-threat approach to the experience of stereotype threats. *Personality and Social Psychology Bulletin*, 37(4), 464-480. doi:10.1177/0146167211398140
- Shapiro, J. R. (2012). Types of threats: From stereotype threat to stereotype threats. In M. Inzlicht, & T. Schmader (Eds.), *Stereotype threat: Theory, process, and applications* (pp. 71-88) Oxford University Press New York, NY.

- Shapiro, J. R., & Neuberg, S. L. (2007). From stereotype threat to stereotype threats: Implications of a multi-threat framework for causes, moderators, mediators, consequences, and interventions. *Personality and Social Psychology Review*, 11(2), 107-130. doi:10.1177/1088868306294790
- Shapiro, J. R., Williams, A. M., & Hambarchyan, M. (2013). Are all interventions created equal? A multi-threat approach to tailoring stereotype threat interventions. *Journal of Personality and Social Psychology*, 104(2), 277. doi:10.1037/a0030461
- Shih, M., Pittinsky, T. L., & Ambady, N. (1999). Stereotype susceptibility: Identity salience and shifts in quantitative performance. *Psychological Science*, 10(1), 80-83. doi:10.1111/1467-9280.00111
- Siegle, D., & Reis, S. M. (1998). Gender differences in teacher and student perceptions of gifted students' ability and effort. *Gifted Child Quarterly*, 42(1), 39-47. doi:10.1177/001698629804200105
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22(11), 1359-1366. doi:10.1177/0956797611417632
- Skaalvik, S., & Skaalvik, E. M. (2004). Gender differences in math and verbal self-concept, performance expectations, and motivation. *Sex Roles*, 50(3-4), 241-252.
- Smith, C. E., & Hopkins, R. (2004). Mitigating the impact of stereotypes on academic performance: The effects of cultural identity and attributions for success among african american college students. *Western Journal of Black Studies*, 28(1), 312.
- Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology*, 35(1), 4-28. doi:10.1006/jesp.1998.1373
- Sroufe, L. A. (1985). Attachment classification from the perspective of infant-caregiver relationships and infant temperament. *Child Development*, , 1-14. doi:10.2307/1130168
- Stanley, J. C. (1971). Predicting college success of the educationally disadvantaged. *Science*, 171(3972), 640-647.
- Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, 52(6), 613. doi:10.1037/0003-066X.52.6.613
- Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of african americans. *Journal of Personality and Social Psychology*, 69(5), 797. doi:10.1037/0022-3514.69.5.797

- Steele, C. M., Spencer, S. J., & Aronson, J. (2002). Contending with group image: The psychology of stereotype and social identity threat. *Advances in Experimental Social Psychology*, 34, 379-440. doi:10.1016/S0065-2601(02)80009-0
- Stockard, J. (1999). Gender socialization. In S. J. Chafetz (Ed.), *Handbook of the sociology of gender* (pp. 215-227) Springer.
- Stoet, G., & Geary, D. C. (2012). Can stereotype threat explain the gender gap in mathematics performance and achievement? *Review of General Psychology*, 16(1), 93. doi:10.1037/a0026617
- Stoet, G., & Geary, D. C. (2018). The gender-equality paradox in science, technology, engineering, and mathematics education. *Psychological Science*, 29(4), 581-593. doi:10.1177/0956797617741719
- Stone, J., Lynch, C. I., Sjomeling, M., & Darley, J. M. (1999). Stereotype threat effects on black and white athletic performance. *Journal of Personality and Social Psychology*, 77(6), 1213. doi:10.1037/0022-3514.77.6.1213
- Stromquist, N. P. (2007). The gender socialization process in schools: A cross-national comparison. *Background Paper Prepared for the Education for all Global Monitoring Report 2008*,
- Tajfel, H. (1978). Interindividual behaviour and intergroup behaviour. *Differentiation between Social Groups: Studies in the Social Psychology of Intergroup Relations*, , 27-60.
- Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. *The Social Psychology of Intergroup Relations*, 33(47), 74.
- Tante, A. C. (2018). 'I can talk about a lot of things in the other language but not in english': Teaching speaking skills in cameroon primary schools. In A. Burns, & J. Siegel (Eds.), *International perspectives on teaching the four skills in ELT* (pp. 81-95) Springer.
- Tashakkori, A., & Teddlie, C. (1998). *Mixed methodology: Combining qualitative and quantitative approaches* Sage.
- Tashakkori, A., & Teddlie, C. (2010). *Sage handbook of mixed methods in social & behavioral research* Sage.
- Tavakoli, H. (2013). *A Dictionary of Language Acquisition: A Comprehensive Overview of Key Terms in First and Second Language Acquisition*. Tehran: Rahnama Press.
- Teddlie, C., & Tashakkori, A. (2003). *Handbook of mixed methods in social & behavioral research* Sage.

- Teddlie, C., & Tashakkori, A. (2009). *Foundations of mixed methods research: Integrating quantitative and qualitative approaches in the social and behavioral sciences* Sage.
- Teddlie, C., & Tashakkori, A. (2012). Common "core" characteristics of mixed methods research: A review of critical issues and call for greater convergence. *American Behavioral Scientist*, 56(6), 774-788. doi:10.1177/0002764211433795
- Teimouri, Y., Goetze, J., & Plonsky, L. (2019). Second language anxiety and achievement: A meta-analysis. *Studies in Second Language Acquisition*, 1-25. doi: 10.1017/S0272263118000311
- Temple, B., & Young, A. (2004). Qualitative research and translation dilemmas. *Qualitative Research*, 4(2), 161-178. doi:10.1177/1468794104044430
- Thompson, W. E., Hickey, J. V., & Thompson, M. L. (2016). *Society in focus: An introduction to sociology* Rowman & Littlefield.
- Thornbury, S. (2012). Speaking instruction. In A. Burns, & J. C. Richards (Eds.), *The cambridge guide to pedagogy and practice in second language teaching* (pp. 198-206). Cambridge: New York: Cambridge University Press.
- Tiedemann, J. (2002). Teachers' gender stereotypes as determinants of teacher perceptions in elementary school mathematics. *Educational Studies in Mathematics*, 50(1), 49-62. doi:1020518104346
- Truitt, S. N. (1996). Anxiety and beliefs about language learning: A study of korean university students learning english. *Texas Papers in Foreign Language Education*, 2(1), n1.
- Tsai, P., & Cheng, Y. (2009). The effects of rhetorical task type, english proficiency, and writing anxiety on senior high school students. *English Teaching & Learning*, 33(3), 95-131. doi:10.6330/ETL.2009.33.3.03
- Tsui, A. B. M. (1996). Reticence and anxiety in second language learning. In K. M. Bailey, & D. Nunan (Eds.), *Voices from the language classroom* (pp. 145-167). New York: Cambridge University Press.
- Tsui, A. B., & Tollefson, J. W. (Eds.). (2007). *Language policy and the construction of national cultural identity* Routledge.
- Turkish Statistical Institute. (2018). *Migration statistics*. Retrieved from http://www.turkstat.gov.tr/VeriBilgi.do?alt_id=1067
- Twinn, D. S. (1998). An analysis of the effectiveness of focus groups as a method of qualitative data collection with chinese populations in nursing research. *Journal of Advanced Nursing*, 28(3), 654-661. doi:0.1046/j.1365-2648.1998.00708.x

- Uğurlu, N. S., Türkoğlu, B., Kuzlak, A., & Gupta, A. (2018). Stereotypes of single and married women and men in turkish culture. *Current Psychology*, 1-13. doi:10.1007/s12144-018-9920-9
- Vale, D., Ozen, E. N., & Alpaslan, I. B. (2013). *Turkey national needs assessment of state school english language teaching*. Ankara: British Council and TEPAV.
- van der Vleuten, M., Jaspers, E., Maas, I., & van der Lippe, T. (2016). Boys' and girls' educational choices in secondary education. the role of gender ideology. *Educational Studies*, 42(2), 181-200. doi:10.1080/03055698.2016.1160821
- Van Teijlingen, E. R., & Hundley, V. (2001). The importance of pilot studies. *Social Research Update*, (35)
- Van Teijlingen, E. R., & Hundley, V. (2002). The importance of pilot studies. *Nursing Standard*, 16(40), 33-36. doi:10.7748/ns2002.06.16.40.33.c3214
- Vandergrift, L. (2007). Recent developments in second and foreign language listening comprehension research. *Language Teaching*, 40(3), 191-210. doi:10.1017/S0261444807004338
- Vandergrift, L. (1997). The comprehension strategies of second language (french) listeners: A descriptive study. *Foreign Language Annals*, 30(3), 387-409.
- VanPatten, B., & Benati, A. G. (2015). *Key terms in second language acquisition* Bloomsbury Publishing.
- Vogely, A. J. (1998). Listening comprehension anxiety: Students' reported sources and solutions. *Foreign Language Annals*, 31(1), 67-80. doi:10.1111/j.1944-9720.1998.tb01333.x
- Vraštilová, O. (2018). Supporting elementary students' reading through authentic literature for children. In A. Burns, & J. Siegel (Eds.), *International perspectives on teaching the four skills in ELT* (pp. 141-152) Springer.
- Vygotsky, L. S. (1986). *Thought and language*. Cambridge, MA: The MIT Press.
- Wang, C., Kim, D., Bai, R., & Hu, J. (2014). Psychometric properties of a self-efficacy scale for english language learners in china. *System*, 44, 24-33. doi:10.1016/j.system.2014.01.015
- Wang, C., Kim, D., Bong, M., & Ahn, H. S. (2013). Examining measurement properties of an english self-efficacy scale for english language learners in korea. *International Journal of Educational Research*, 59, 24-34. doi:10.1016/j.ijer.2013.02.004
- West, R., Güven, A., & Ergenekon, T. (2015). *The state of english in higher education in turkey. A baseline study*. Ankara: Yorum Press.

- Wharton, A. S. (2005). *The sociology of gender: An introduction to theory and research*. John Wiley & Sons.
- Wheelan, S. A. (2005). *The handbook of group research and practice* Sage.
- Whitehead, J. M. (1996). Sex stereotypes, gender identity and subject choice at a-level. *Educational Research*, 38(2), 147-160. doi:10.1080/0013188960380203
- Widaman, K. F., Little, T. D., Preacher, K. J., & Sawalani, G. M. (2011). On creating and using short forms of scales in secondary research. doi:10.1037/12350-003
- Willig, C. (2013). *Introducing qualitative research in psychology* McGraw-Hill Education (UK).
- Wilson, N., & McClean, S. I. (1994). *Questionnaire design: A practical introduction* University of Ulster Newtownabbey.
- Woodrow, L. (2006). Anxiety and speaking english as a second language. *RELC Journal*, 37(3), 308-328. doi:10.1177/0033688206071315
- Wout, D., Danso, H., Jackson, J., & Spencer, S. (2008). The many faces of stereotype threat: Group-and self-threat. *Journal of Experimental Social Psychology*, 44(3), 792-799. doi:10.1016/j.jesp.2007.07.005
- Wucherer, B. V., & Reiterer, S. M. (2016). Language is a girly thing, isn't it? A psycholinguistic exploration of the L2 gender gap. *International Journal of Bilingual Education and Bilingualism*, 21(1), 118-134. doi:10.1080/13670050.2016.1142499
- Yamashita, J. (2004). Reading attitudes in L1 and L2, and their influence on L2 extensive reading. *Reading in a Foreign Language*, 16(1), 1-19.
- Yamashita, J. (2007). The relationship of reading attitudes between L1 and L2: An investigation of adult EFL learners in japan. *TESOL Quarterly*, 41(1), 81-105. doi: 10.1002/j.1545-7249.2007.tb00041.x
- Yvonne Feilzer, M. (2010). Doing mixed methods research pragmatically: Implications for the rediscovery of pragmatism as a research paradigm. *Journal of Mixed Methods Research*, 4(1), 6-16. doi:10.1177/1558689809349691
- Zhang, X. (2013). Foreign language listening anxiety and listening performance: Conceptualizations and causal relationships. *System*, 41(1), 164-177. doi:10.1016/j.system.2013.01.004
- Zhao, A., Guo, Y., & Dynia, J. (2013). Foreign language reading anxiety: Chinese as a foreign language in the united states. *The Modern Language Journal*, 97(3), 764-778. doi:10.1111/j.1540-4781.2013.12032.x

- Zimmerman, B. J., & Cleary, T. J. (2006). Adolescents' development of personal agency: The role of self-efficacy beliefs and self-regulatory skill. *Self-Efficacy Beliefs of Adolescents*, 5, 45-69.
- Zimmerman, B. J., Schunk, D. H., & DiBenedetto, M. K. (2017). The role of Self-Efficacy and related beliefs in Self-Regulation of learning and performance. In A. J. Elliot, C. S. Dweck & D. S. Yeager (Eds.), *Handbook of competence and motivation: Theory and application* (pp. 313-333) Guilford Publications.
- Zin, M. (2010). Anxiety and academic reading performance among malay ESL learners. *Journal of Pan-Pacific Association of Applied Linguistics*, 14(2), 41-58.
- Ziya, H. E., & Koç, Y. (2016). Being a gay man in turkey: Internalised sexual prejudice as a function of prevalent hegemonic masculinity perceptions. *Culture, Health & Sexuality*, 18(7), 799-811. doi:10.1080/13691058.2015.1133846

APPENDICES

Appendix 3.1. Ethical Approval Letter

RE: ethics application - The Effect of Stereotype Threat on Foreign Language Performance through the Mediating roles of Self-Efficacy Beliefs and Language Anxiety
2 messages

Julie Kirby <Kirbyjul@edgehill.ac.uk>
To: Gulsah Kutuk <gulsah.kutuk@go.edgehill.ac.uk>
Cc: Tim Cain <Caint@edgehill.ac.uk>

26 September 2016 at 13:03

Dear Gulsah,

I am pleased to confirm that your third study has been given ethical approval, after being re-considered by members of FREC. This means that your entire study now has ethical approval.

Please retain this email for your records.

Kind regards

Julie

Julie Kirby
Faculty Research Administrator
Edge Hill University
FoE Annex Building
St Helens Road
Ormskirk
L39 4QP
Tel: 01695 584235
Email: kirbyjul@edgehill.ac.uk

[Edge Hill University](#)
[Times Higher Education University of the Year](#)

Appendix 3.2. Participant Information Sheets

STUDY 1

PARTICIPANT INFORMATION SHEET

- **Invitation Paragraph**

I am a PhD student conducting some research examining the psychological processes underpinning learning a new language. I will really appreciate your time and efforts in completing the following three questionnaires. However, you should only participate if you want to; choosing not to take part will not disadvantage you in anyway. Before you decide whether you want to take part, it is important for you to understand what your participation will involve. Please take time to read the following information carefully and ask me if there is anything that is not clear or if you would like more information.

- **Why have I been invited to take part?**

I would like to invite you to participate in this research project because you are a student learning English as a foreign language.

- **Do I have to take part?**

Participation is voluntary in this study. You also have the right to withdraw from this study if you change your mind at any point. This can be simply done by informing me either during the study or after the study through the email given below. You do not need to have a reason to withdraw from the study, but you will not need to submit your request within 4 weeks of participating. When you request to withdraw from the study, all the data you have provided will be destroyed.

- **What will happen to me if I take part?**

If you decide to take part, you will be given this information sheet to keep and will be asked to sign a consent form. Following this, you will be given three different questionnaires to be filled in one go.

You will also be asked if you would like to participate in Study 2. Should you wish to participate Study 2 as well, I will then contact you to discuss the interview procedure with you at a time convenient for you. On request you will be given the interview topic guide. With your consent, I will arrange to interview you in a private area on the premises where you study (or at a suitable venue in a local public site if you prefer).

- **Incentives**

There is no financial incentive to participate in this research. There will be no expenses to spend, either. The study will be conducted in your university.

- **What are the possible risks of taking part?**

There are no foreseeable risks in participating in the study. The main disadvantage to taking part in the study is that you will be donating around an hour of your time to take part.

- **What are the possible benefits of taking part?**

There are no direct benefits to taking part. However, the information I get from the study will help to influence current and future teaching practices in the classrooms.

- **Will my taking part be kept confidential?**

All the data provided in the questionnaires is regarded as strictly confidential and will be held securely until the research is finished. All data for analysis will be anonymised unless permission to use your name is given. In reporting on the research findings, I will not reveal the names of any participants or the organisation where you study. If no permission is given, at all times there will be no possibility of you as an individual being linked with the data.

- **What will happen to the results of the study?**

I will produce a final report summarising the main findings. These findings will be presented in my PhD dissertation. The results will also be disseminated through papers, articles etc.

- **What if something goes wrong?**

If any further information is required or you wish to contact us, please use the following email addresses: Gulsah Kutuk, Kutukg@edgehill.ac.uk

STUDY 2

PARTICIPANT INFORMATION SHEET

- **Invitation Paragraph**

I am a PhD student conducting some research examining the psychological processes underpinning learning a new language. I will really appreciate your time and efforts in answering my interview questions. However, you should only participate if you want to; choosing not to take part will not disadvantage you in anyway. Before you decide whether you want to take part, it is important for you to understand what your participation will involve. Please take time to read the following information carefully and ask the researcher if there is anything that is not clear or if you would like more information.

- **Why have I been invited to take part?**

I would like to invite you to participate in this research project because you teach/learn English as a foreign language.

- **Do I have to take part?**

Participation is voluntary in this study. You also have the right to withdraw from this study if you change your mind at any point. This can be simply done by informing the researcher either during the study or after the study through the email given below. You do not need to have a reason to withdraw from the study, but you will not need to submit your request within 4 weeks of participating. When you request to withdraw from the study, all the data you have provided will be destroyed.

- **What will happen to me if I take part?**

If you decide to take part, you will be given this information sheet to keep and will be asked to sign a consent form. Following this, I will then call you to discuss the interview procedure with you at a time convenient for you. On request you will be given the interview topic guide. With your consent, I will arrange to interview you in a private area on the premises where you study (or at a suitable venue in a local public site if you prefer).

The interview will take approximately an hour and be based on the interview topic guide, but it is designed to be flexible so as to meet your needs. The interview will be

recorded, subject to your permission. Recordings of interviews will be deleted after transcription. Even if you have decided to take part, you are still free to cease your participation at any time and to have research data/information relating to you withdrawn without giving any reason up to the point of transcription at the end of August, 2017.

- **Incentives**

There is no financial incentive to participate in this research. There will be no expenses to spend, either. The study will be conducted in your university.

- **What are the possible risks of taking part?**

There are no foreseeable risks in participating in the study. The main disadvantage to taking part in the study is that you will be donating around an hour of your time to take part.

- **What are the possible benefits of taking part?**

There are no direct benefits to taking part. However, the information I get from the study will help to influence current and future teaching practices in the classrooms.

- **Will my taking part be kept confidential?**

All the information given during the interviews is regarded as strictly confidential and will be held securely until the research is finished. Your participation is entirely voluntary. If you change your mind, you are free to stop your participation and to have your data withdrawn without giving any reason up to the point of transcription at the end of July, 2015. All data for analysis will be anonymised unless permission to use your name is given. In reporting on the research findings, I will not reveal the names of any participants or the organisation where you work. If no permission is given, at all times there will be no possibility of you as an individual being linked with the data.

The UK Data Protection Act 1998 will apply to all information gathered within the interviews and held on password-locked computer files. No data will be accessed by anyone other than me; and anonymity of the material will be protected by using false names. You may withdraw your data from the project anytime up to the point of transcription at the end of August, 2017. All recordings of data on audio-equipment will be deleted after transcription. If you ask me to withdraw your data at any time before August, 2017 I will remove all traces of it from the records.

- **What will happen to the results of the study?**

I will produce a final report summarising the main findings. These findings will be presented in my PhD dissertation. The results will also be disseminated through papers, articles etc.

- **What if something goes wrong?**

If any further information is required or you wish to contact us, please use the following email addresses: Gulsah Kutuk, Kutukg@edgehill.ac.uk

STUDY 3

PARTICIPANT INFORMATION SHEET

- **Invitation Paragraph**

I am a PhD student conducting some research examining the psychological processes underpinning learning a new language. I will really appreciate your time and

efforts in answering my interview questions. However, you should only participate if you want to; choosing not to take part will not disadvantage you in anyway. Before you decide whether you want to take part, it is important for you to understand what your participation will involve. Please take time to read the following information carefully and ask the researcher if there is anything that is not clear or if you would like more information.

- **Why have I been invited to take part?**

I would like to invite you to participate in this research project because you are a student learning English as a foreign language.

- **Do I have to take part?**

Participation is voluntary in this study. You also have the right to withdraw from this study if you change your mind at any point. This can be simply done by informing the researcher either during the study or after the study through the email given below. You do not need to have a reason to withdraw from the study, but you will not need to submit your request within 4 weeks of participating. When you request to withdraw from the study, all the data you have provided will be destroyed.

- **What will happen to me if I take part?**

If you decide to take part, you will be given this information sheet to keep and will be asked to sign a consent form. Following this, you will be given some instructions. Based on these instructions, you will be asked to complete two questionnaires.

On the completion of two questionnaires, I will assess your speaking ability in English. For this assessment, you will be asked some questions in English and you will be required to answer these questions in English. This will not take longer than 10 minutes.

- **Incentives**

There is no financial incentive to participate in this research. There will be no expenses to spend, either. The study will be conducted in your university.

- **What are the possible risks of taking part?**

There are no foreseeable risks in participating in the study. The main disadvantage to taking part in the study is that you will be donating around an hour of your time to take part.

- **What are the possible benefits of taking part?**

There are no direct benefits to taking part. However, the information I get from the study will help to influence current and future teaching practices in the classrooms.

- **Will my taking part be kept confidential?**

All the data provided in the questionnaires is regarded as strictly confidential and will be held securely until the research is finished. All data for analysis will be anonymised unless permission to use your name is given. In reporting on the research findings, I will not reveal the names of any participants or the organisation where you study. If no permission is given, at all times there will be no possibility of you as an individual being linked with the data.

- **What will happen to the results of the study?**

I will produce a final report summarising the main findings. These findings will be presented in my PhD dissertation. The results will also be disseminated through papers, articles etc.

- **What if something goes wrong?**

If any further information is required or you wish to contact us, please use the following email addresses: Gulsah Kutuk, Kutukg@edgehill.ac.uk

Appendix 3.3. Participant Consent Forms

STUDY 1

CONSENT FORM

Before taking part in this research please read the statements below, to indicate you have read and understood the requirements of this study.

- ☐ I have read and understood Participant Brief Form presenting the requirements of the study.
- ☐ I have been informed that my information will be kept confidential and will not be shared with third parties.
- ☐ I understand that the information I give will be used for academic purposes and I consent to my contribution being used in this way.
- ☐ I understand that I can withdraw from this study at any time. I understand this can be done during the process of completing the study by informing the researcher and up to four weeks from participation by contacting the researcher via the email address provided by the researcher.
- ☐ I understand that as part of this study, the researcher will need the exam grades of participants. By giving consent I allow the researcher to obtain my academic record of achievement.
- ☐ I confirm that I am happy to proceed with this study

Sign:

Date:

STUDY 2 AND 3

CONSENT FORM

Before taking part in this research please read the statements below, to indicate you have read and understood the requirements of this study.

- ☐ I have read and understood Participant Brief Form presenting the requirements of the study.
- ☐ I have been informed that my information will be kept confidential and will not be shared with third parties.
- ☐ I understand that the information I give will be used for academic purposes and I consent to my contribution being used in this way.

☐ I understand that I can withdraw from this study at any time. I understand this can be done during the process of completing the study by informing the researcher and up to four weeks from participation by contacting the researcher via the email address provided by the researcher.

☐ I confirm that I am happy to proceed with this study

Sign:

Date:

Appendix 3.4: Participant Debrief Forms

STUDY 1

PARTICIPANT DEBRIEF

The Effect of Stereotype Threat on Foreign Language Performance through the Mediating Roles of Self-Efficacy and Language Anxiety

Thank you for your time in helping us with our research. The aim of this study is to assess whether there is a link between your gender stereotype endorsement level and your anxiety levels, self-efficacy beliefs and language learning performance. In particular, I am most interested in how this may be relevant for male learners of English.

As part of research, your exam results will be evaluated by me. If you do not want me to use your exam results in this research, please contact me via the email given below.

Please be assured that all data obtained from this study will be anonymous and will be kept confidential and be used only for academic purposes. Additionally, you also have the right to withdraw from this study within four weeks of participating, by contacting me via the below contact details.

If you have any further questions regarding any aspect of this study, please do not hesitate to ask me.

Gulsah Kutuk
Kutukg@edgehill.ac.uk

STUDY 2

PARTICIPANT DEBRIEF

The Effect of Stereotype Threat on Foreign Language Performance through the Mediating Roles of Self-Efficacy and Language Anxiety

Thank you for your time in helping us with our research. The aim of this study is to assess whether there is a link between your gender stereotyped beliefs about language learning and your teaching practices in the classrooms. In particular, I am most interested in how you treat male learners of English.

Please be assured that all data obtained from this study will be anonymous, and will be kept confidential and be used only for academic purposes. Additionally, you also have

the right to withdraw from this study within four weeks of participating, by contacting me via the below contact details.

If you have any further questions regarding any aspect of this study, please do not hesitate to ask me.

Gulsah Kutuk
Kutukg@edgehill.ac.uk

STUDY 3

PARTICIPANT DEBRIEF

The Effect of Stereotype Threat on Foreign Language Performance through the Mediating Roles of Self-Efficacy and Language Anxiety

Thank you for your time in helping us with our research. It is important to state that any negative stereotypes you have heard are not a true reflection of your language ability and were only used to examine the effects that negative stereotypes may have on language performance.

The aim of this study is to assess whether an imposed stereotype threat of language learning (e.g., "there is a common stereotype that language learning is a feminine domain") affects actual language learning performance. In particular, I am most interested in how this may be relevant for male learners of English.

Please be assured that all data obtained from this study will be anonymous, and will be kept confidential and be used only for academic purposes. Additionally, you also have the right to withdraw from this study within four weeks of participating, by contacting me via the email given below.

If you have any further questions regarding any aspect of this study, please do not hesitate to ask me.

Gulsah Kutuk
Kutukg@edgehill.ac.uk

Appendix 4.1. Pilot Study Materials

Appendix 4.1.1. Part 1. Background Questionnaire

BACKGROUND INFORMATION

1. Your name and surname:		
2. Your Gender:	<input type="checkbox"/> Female	<input type="checkbox"/> Male
3. Age:		

4. Which part of Turkey are you from?	
5. Where did you study at high school?	
6. What is your mother's job?	
7. What is your father's job?	
8. What is the highest level of education completed by your mother?	
9. What is the highest level of education completed by your father?	
10. Can any of your family members speak English? If yes, who can speak English? Level?	
11. What is your English level?	
12. How long have you been learning English at university?	
13. Have you studied English before university?	
14. How long?	
15. Do you speak any other languages? Which language? Level?	
16. Have you been abroad before?	
17. Which department are you in? (<i>Engineering, medicine, social sciences etc.</i>)	

Appendix 4.1.2. Part 2: Learners' Gender Stereotyped *Perceptions of Subjects*

LEARNERS' PERCEPTIONS OF ACADEMIC SUBJECTS

Please put a (√) in the appropriate boxes.

Subject	1 Female s better at:	2	3 Both sexes equally good at:	4	5 Males better at:
1. Social Sciences, e.g. Sociology, Psychology					
2. Biological Sciences, e.g. Chemistry, Biology					

3. Performing Arts, e.g. Art, Music, Drama					
4. Geography and related subjects, e.g. Geography					
5. Arts subjects, e.g. English Literature, History					
6. Practical subjects, e.g. Nursing, Office Practice					
7. Languages, e.g. English					
8. Economics and related subjects, e.g. Politics					
9. Mathematics and related subjects, e.g. Statistics					
10. Physical Sciences, e.g. Physics, Chemistry					
11. Engineering, e.g. civil or mechanical engineering					

Appendix 4.1.3. Part 3: *Learners' Perceptions of L2 learning*

LEARNERS' PERCEPTIONS OF FOREIGN LANGUAGE LEARNING

Please put a (√) in the appropriate boxes.

	Strongly Disagree 1	Disagree 2	Neither 3	Agree 4	Strongly Agree 5
1. I think language arts is a feminine domain.					
2. I think females have more ability in language arts than males.					
3. In general, females are better than males at language arts.					
4. I don't think there are any gender differences in language arts ability.					
5. According to my lecturers, females perform better than males in the classroom.					
6. My lecturers think that language learning is a feminine domain.					
7. My lecturers think that females have more ability in language arts than males.					
8. I don't think my lecturers find females more successful than males in the classroom.					

Appendix 4.2. Multidimensional Language Class Anxiety Scale (MCLAS)
SKILL-BASED FOREIGN LANGUAGE LEARNING ANXIETY SCALE

1	2	3	4	5
Strongly Disagree	Disagree	Neither Agree nor disagree	Agree	Strongly Agree

PART 1:

READING CLASS ANXIETY

1. Thinking about reading activities in class makes me feel uneasy.	1	2	3	4	5
2. I feel nervous during reading activities in class.	1	2	3	4	5
3. Even before reading activities in class, I worry whether I will be able to understand the material.	1	2	3	4	5
4. I worry whether I am sufficiently prepared for reading activities in class	1	2	3	4	5
5. I get tense during reading activities in class.	1	2	3	4	5
6. When I don't understand something important during reading activities in class, my heart races.	1	2	3	4	5

PART 2:

WRITING CLASS ANXIETY

7. Thinking about writing activities in class makes me feel uneasy.	1	2	3	4	5
8. I feel nervous during writing activities in class.	1	2	3	4	5
9. Even before writing activities in class, I worry whether I will be able to understand the material.	1	2	3	4	5
10. I worry whether I am sufficiently prepared for writing activities in class	1	2	3	4	5
11. I get tense during writing activities in class.	1	2	3	4	5
12. When I don't understand something important during writing activities in class, my heart races.	1	2	3	4	5

PART 3:

LISTENING CLASS ANXIETY

13. Thinking about listening activities in class makes me feel uneasy.	1	2	3	4	5
14. I feel nervous during listening activities in class.	1	2	3	4	5

15. Even before listening activities in class, I worry whether I will be able to understand the material.	1	2	3	4	5
16. I worry whether I am sufficiently prepared for listening activities in class	1	2	3	4	5
17. I get tense during listening activities in class.	1	2	3	4	5
18. When I don't understand something important during listening activities in class, my heart races.	1	2	3	4	5

PART 4

SPEAKING CLASS ANXIETY

19. Thinking about speaking activities in class makes me feel uneasy.	1	2	3	4	5
20. I feel nervous during speaking activities in class.	1	2	3	4	5
21. Even before speaking activities in class, I worry whether I will be able to understand the material.	1	2	3	4	5
22. I worry whether I am sufficiently prepared for speaking activities in class.	1	2	3	4	5
23. I get tense during speaking activities in class.	1	2	3	4	5
24. When I don't understand something important during speaking activities in class, my heart races.	1	2	3	4	5

PART 5

TESTING-RELATED ANXIETY

25. Thinking about taking an English language test makes me feel uneasy.	1	2	3	4	5
26. I feel nervous while taking an English language test.	1	2	3	4	5
27. Even before taking an English language test, I worry whether I will be able to understand the material.	1	2	3	4	5
28. I worry whether I am sufficiently prepared for taking an English language test.	1	2	3	4	5
29. I get tense while taking an English language test.	1	2	3	4	5
30. When I don't understand something important while taking English language test, my heart races.	1	2	3	4	5

Appendix 4.3. Common Reference Levels: self-assessment grid

		A1	A2
U N D E R S T A N D I N G	Listening	I can recognise familiar words and very basic phrases concerning myself, my family and immediate concrete surroundings when people speak slowly and clearly.	I can understand phrases and the highest frequency vocabulary related to areas of most immediate personal relevance (e.g. very basic personal and family information, shopping, local area, employment). I can catch the main point in short, clear, simple messages and announcements.
	Reading	I can understand familiar names, words and very simple sentences, for example on notices and posters or in catalogues.	I can read very short, simple texts. I can find specific, predictable information in simple everyday material such as advertisements, prospectuses, menus and timetables and I can understand short simple personal letters.
S P E A K I N G	Spoken Interaction	I can interact in a simple way provided the other person is prepared to repeat or rephrase things at a slower rate of speech and help me formulate what I'm trying to say. I can ask and answer simple questions in areas of immediate need or on very familiar topics.	I can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar topics and activities. I can handle very short social exchanges, even though I can't usually understand enough to keep the conversation going myself.
	Spoken Production	I can use simple phrases and sentences to describe where I live and people I know.	I can use a series of phrases and sentences to describe in simple terms my family and other people, living conditions, my educational background and my present or most recent job.
W R I T I N G	Writing	I can write a short, simple postcard, for example sending holiday greetings. I can fill in forms with personal details, for example entering my name, nationality and address on a hotel registration form.	I can write short, simple notes and messages relating to matters in areas of immediate needs. I can write a very simple personal letter, for example thanking someone for something.

		B1	B2
U N D E R S T A N D I N G	Listening	I can understand the main points of clear standard speech on familiar matters regularly encountered in work, school, leisure, etc. I can understand the main point of many radio or TV programmes on current affairs or topics of personal or professional interest when the delivery is relatively slow and clear.	I can understand extended speech and lectures and follow even complex lines of argument provided the topic is reasonably familiar. I can understand most TV news and current affairs programmes. I can understand the majority of films in standard dialect.
	Reading	I can understand texts that consist mainly of high frequency everyday or job-related language. I can understand the description of events, feelings and wishes in personal letters.	I can read articles and reports concerned with contemporary problems in which the writers adopt particular attitudes or viewpoints. I can understand contemporary literary prose.
S P E A K I N G	Spoken Interaction	I can deal with most situations likely to arise whilst travelling in an area where the language is spoken. I can enter unprepared into conversation on topics that are familiar, of personal interest or pertinent to everyday life (e.g. family, hobbies, work, travel and current events).	I can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible. I can take an active part in discussion in familiar contexts, accounting for and sustaining my views.
	Spoken Production	I can connect phrases in a simple way in order to describe experiences and events, my dreams, hopes and ambitions. I can briefly give reasons and explanations for opinions and plans. I can narrate a story or relate the plot of a book or film and describe my reactions.	I can present clear, detailed descriptions on a wide range of subjects related to my field of interest. I can explain a viewpoint on a topical issue giving the advantages and disadvantages of various options.
W R I T I N G	Writing	I can write simple connected text on topics which are familiar or of personal interest. I can write personal letters describing experiences and impressions.	I can write clear, detailed text on a wide range of subjects related to my interests. I can write an essay or report, passing on information or giving reasons in support of or against a particular point of view. I can write letters highlighting the personal significance of events and experiences.

		C1	C2
U N D E R S T A N D I N G	Listening	I can understand extended speech even when it is not clearly structured and when relationships are only implied and not signalled explicitly. I can understand television programmes and films without too much effort.	I have no difficulty in understanding any kind of spoken language, whether live or broadcast, even when delivered at fast native speed, provided I have some time to get familiar with the accent.
	Reading	I can understand long and complex factual and literary texts, appreciating distinctions of style. I can understand specialised articles and longer technical instructions, even when they do not relate to my field.	I can read with ease virtually all forms of the written language, including abstract, structurally or linguistically complex texts such as manuals, specialised articles and literary works.
S P E A K I N G	Spoken Interaction	I can express myself fluently and spontaneously without much obvious searching for expressions. I can use language flexibly and effectively for social and professional purposes. I can formulate ideas and opinions with precision and relate my contribution skilfully to those of other speakers.	I can take part effortlessly in any conversation or discussion and have a good familiarity with idiomatic expressions and colloquialisms. I can express myself fluently and convey finer shades of meaning precisely. If I do have a problem I can backtrack and restructure around the difficulty so smoothly that other people are hardly aware of it.
	Spoken Production	I can present clear, detailed descriptions of complex subjects integrating sub-themes, developing particular points and rounding off with an appropriate conclusion.	I can present a clear, smoothly-flowing description or argument in a style appropriate to the context and with an effective logical structure which helps the recipient to notice and remember significant points.
W R I T I N G	Writing	I can express myself in clear, well-structured text, expressing points of view at some length. I can write about complex subjects in a letter, an essay or a report, underlining what I consider to be the salient issues. I can select style appropriate to the reader in mind.	I can write clear, smoothly-flowing text in an appropriate style. I can write complex letters, reports or articles which present a case with an effective logical structure which helps the recipient to notice and remember significant points. I can write summaries and reviews of professional or literary works.

Appendix 4.4. Questionnaire of Self-Efficacy Beliefs in Learning a New Language (QSLL)

QUESTIONNAIRE OF SELF-EFFICACY BELIEFS IN LEARNING A NEW LANGUAGE

1	2	3	4	5
Strongly Disagree	Disagree	Neither Agree nor disagree	Agree	Strongly Agree

PART 1:

LISTENING EFFICACY:

1. I can understand familiar everyday expressions and very basic phrases in an audio-recorded English text.	1	2	3	4	5
2. I can understand someone speaking about himself/his family and friends in English.	1	2	3	4	5
3. I can understand the main point of an English radio/TV program on a personal /professional interest.	1	2	3	4	5
4. I can understand English TV news programs without English/Turkish subtitles.	1	2	3	4	5
5. I can understand English films without English/Turkish subtitles.	1	2	3	4	5

PART 2:

SPEAKING EFFICACY:

6. I can use simple English phrases and sentences to describe where I live and people I know.	1	2	3	4	5
7. I can ask questions to my teacher and answer his/her questions in English.	1	2	3	4	5
8. I can discuss topics such as families, hobbies, work and travel with my classmates in English.	1	2	3	4	5
9. I can interact with a native speaker of English fluently and spontaneously.	1	2	3	4	5
10. I can express myself fluently and spontaneously without much obvious searching for expressions in English.	1	2	3	4	5

PART 3:

READING EFFICACY:

11. I can read and understand very simple English sentences on notices, posters or in catalogues.	1	2	3	4	5
---	---	---	---	---	---

12. I can read and understand very short, simple texts such as English graded readers.	1	2	3	4	5
13. I can read and understand a personal letter describing events, feelings and wishes in English.	1	2	3	4	5
14. I can read and understand English articles and reports concerned with contemporary problems.	1	2	3	4	5
15. I can read and understand long and complex factual and literary English texts (e.g., novels, articles, essays etc.).	1	2	3	4	5

PART 4:

WRITING EFFICACY:					
16. I can write a short, simple postcard to my friend in English (E.g., sending holiday greetings).	1	2	3	4	5
17. I can write English notes and messages to my friends.	1	2	3	4	5
18. I can write a personal letter describing my experiences and impressions in English.	1	2	3	4	5
19. I can write an English essay giving reasons in support of or against a particular point of view.	1	2	3	4	5
20. I can express myself in clear well-structured English text, expressing points of view at some length.	1	2	3	4	5

Appendix 4.5. Revised Version of Learners' Perceptions About the Academic Subjects

1	2	3	4	5
Females are always successful	Females are often successful	Sometimes females sometimes males are successful	Males are often successful	Males are always successful

ACADEMIC SUBJECTS					
1. Fine arts. For example; Music, Art, Sculpture.	1	2	3	4	5
2. Educational Sciences. For example; Mathematics Education, Preschool Education, Education Management.	1	2	3	4	5
3. Sports Sciences. For Example; Physical and Sport Education, Sport Management, Coaching Education.	1	2	3	4	5

4. Aviation and Space Sciences. For example; Civil Air Transportation Management, Pilotage, Air Traffic Management.	1	2	3	4	5
5. Architecture and Design. For example, Interior Design, Architecture, Fashion Design.	1	2	3	4	5
6. Economics and Administrative Sciences. For example, Business, Finance, Public Administration.	1	2	3	4	5
7. Medical Sciences. For Example, Biophysics, Neurology, Orthopaedics and Traumatology.	1	2	3	4	5
8. Dentistry. For example, Oral and Maxillofacial Surgery, Orthodontics, Endodontics.	1	2	3	4	5
9. Health Sciences. For example, Child Development, Nutrition and Dietetics, Nursing.	1	2	3	4	5
10. Life Sciences. For example, Physics, Chemistry, Mathematics.	1	2	3	4	5
11. Humanities. For example, Literature, History, Philosophy.	1	2	3	4	5
12. Languages. For example, English	1	2	3	4	5
13. Engineering. For example, Computer Engineering, Industrial Engineering, Chemical Engineering.	1	2	3	4	5
14. Communication. For example, Public Relations and Publicity, Radio, Television and Film. Advertising.	1	2	3	4	5
15. Tourism. For example, Tourism Management, Tourist Guiding, Gastronomy and Culinary Arts.	1	2	3	4	5
16. Agriculture. For example, Agricultural Economics, Agricultural Biotechnology, Soil Science and Plant Nutrition	1	2	3	4	5

Appendix 4.6. Revised Version of *Questionnaire of Language Learners' Gender Stereotyped Beliefs*

LEARNERS' PERCEPTIONS ABOUT LEARNING ENGLISH AS A FOREIGN LANGUAGE

1. In this section, you are asked to share your perceptions about learning English as a foreign language. Please read the four statements given below and rate them using

the scale provided. **Don't forget! The section is all about your perceptions. There are no correct or wrong answers.**

1	2	3	4	5
Always females	Often females	Sometimes Females Sometimes Males	Often Males	Always Males

1) Please indicate which gender is generally good at learning English.	1	2	3	4	5
2) Please indicate which gender is eager and motivated to learn English.	1	2	3	4	5
3) Please indicate which gender is good at using the methods and strategies that are effective for learning English.	1	2	3	4	5
4) Please indicate which gender is naturally adept for learning English.	1	2	3	4	5
5) Please indicate which gender performs well in English classes .	1	2	3	4	5

2. In this section, you are asked to share your perceptions about your English instructors. You need to think about the impressions that your English instructors have made on you while teaching English in the classroom. **You don't need to know your English instructors' real thoughts and perceptions about the statements**. Please read the four statements given below and rate them using the scale provided.

1	2	3	4	5
Always Females	Often Females	Sometimes Females Sometimes Males	Often Males	Always Males

6) Please indicate which gender is generally good at learning English according to your English instructors.	1	2	3	4	5
7) Please indicate which gender is eager and motivated to learn English according to your English instructors.	1	2	3	4	5
8) Please indicate which gender is good at using the methods and strategies that are effective for learning English according to your English instructors.	1	2	3	4	5

9) Please indicate which gender is <i>naturally adept</i> for learning English according to your English instructors.	1	2	3	4	5
10) Please indicate which gender <i>performs well in English classes</i> according to your English instructors.	1	2	3	4	5

Appendix 4.7. Revised Version of Questionnaire of Self-Efficacy Beliefs in Learning a New Language (QSLL)

1	2	3	4	5
Strongly Disagree	Disagree	Neither Agree nor disagree	Agree	Strongly Agree

RECEPTION	I can listen to and understand the main point of an English radio/TV program on a personal /professional interest.
	I can watch and understand English TV news programs without English/Turkish subtitles.
	I can watch and understand English films and TV series without English/Turkish subtitles.
	I can read and understand the main point of English articles and reports concerned with contemporary problems without using any kind of dictionaries
	I can read and understand the majority of long and complex English literary texts such as novels and essays without using any kind of dictionaries.
PRODUCTION	I can have a conversation with my classmates and instructors on familiar and daily topics such as families, hobbies, work and travel in English without any preparation in advance.
	During the English class, I can ask questions to my instructors and answer their questions verbally in English.
	I can verbally state my opinions about the contemporary issues or my plans for the future in English.
	I can write a personal letter/an email describing my experiences and impressions in English without using any kind of dictionaries.
	I can write an English essay giving reasons in support of or against a particular point of view without using any kind of dictionaries.
	I can express myself in clear well-structured written English text, expressing points of view at some length without using any kind of dictionaries.

Appendix 5.1. Interview Questions for L2 Teachers

Part 1: Background Information

1. How long have you been teaching English?

2. How did you decide to be an English instructor?
 - a. Why didn't you choose another subject? (E.g. Maths, science, engineering)
3. How did you learn English?
 - a. Why do you think it was easier than the other subjects?

Prompt: Do you think it was because you are talented in language learning?
4. What were the biggest challenges for you as a student?
 - a. Were your language teachers supportive?
 - b. Did you have any issues with them?
5. What are the biggest challenges for you as an English teacher?

Part 2: Beliefs and Perceptions

a. Gender Stereotyping of Language Learning

6. Have you noticed any differences between your female and male students? Which group is more challenging for you?
 - a. Why?
 - b. Any examples?
7. Let's assume that you are not an English teacher. You are a teacher of Maths. What would be different?
 - a. Which group of students would be more challenging?
 - b. Why?

Prompt: So, do think that there are gender differences in subject choices?

 - a. If you have a look at this table, which ones do you think male students/female students/both would be successful at?
 - b. Why?

Subjects

Practical subjects, e.g. Nursing, Office Practice
 Languages, e.g. English
 Performing Arts, e.g. Art, Music, Drama
 Biological Sciences, e.g. Chemistry, Biology
 Arts subjects, e.g. English Literature, History
 Social Sciences, e.g. Sociology, Psychology
 Geography and related subjects, e.g. Geography
 Economics and related subjects, e.g. Politics
 Mathematics and related subjects, e.g. Statistics
 Physical Sciences, e.g. Physics, Chemistry
 Engineering, e.g. civil or mechanical engineering

8. (Focus on language learning) Why do you think it is appropriate for female students/male students/both?

Prompt: Do you think language learning is a feminine domain?

Prompt: Do you think this might be the reason that there are more female English instructors than males?

 - a. Do you think women have more ability in foreign language learning than men?
 - b. Do you think women are better than men at learning a foreign language in general?

b. Males and Self-Efficacy Beliefs

9. Could you please name the most successful three students in your classes?
10. Why do you think these students are successful?
Prompt: What do these students have in common?
11. What do you think about unsuccessful students?
Prompt: What do these students have in common?
Prompt: Do you think confidence in language learning affects their language learning performance?
12. Have you ever seen any differences between male and female students in terms of their confidence in language learning?
 - a. Who do you think is less confident in language learning?
 - b. Why do you think they have low confidence?Prompt: Have you ever had any male students who believe that they can't learn English?
 - c. What were his reasons for that?
 - d. How did you help him?

c. Males and Foreign Language Anxiety

13. Do you think foreign language anxiety affects students' language learning performance?
 - a. How?
 14. Do you believe that there are differences between male and female students in terms of their anxiety levels?
 - a. How do you know that? Any examples?
 15. Who do you think tend to be more anxious when it comes to language learning- males or females? Why?
 - a. Why do you think those students are more anxious during the classes?Prompt: Have you ever seen any differences between male and female students in terms of their anxiety levels?
 16. What do you do to support the students with high level of language learning anxiety?
-

Appendix 5.2. Interview Questions for L2 Learners**Part 1: Background Information**

1. What is your department of study?
2. Why did you choose this program?
Prompt: What was the biggest influence on your decision?
3. How long have you been learning English?
 - a. Do you like learning English?
 - b. What are the biggest challenges for you as a student learning English?
 - c. If you had the option to opt out of language classes, what would you do?
 - d. Why?

Part 2: Beliefs and Perceptions**a. Gender Stereotyping of Language Learning**

4. Do you think you would have the same challenges if you were studying another subject such as Maths?
 - a. Why? Why not?
5. If you had the freedom choose a few subjects to study, which one would you choose?

Subjects

Practical subjects, e.g. Nursing, Office Practice
 Languages, e.g. English
 Performing Arts, e.g. Art, Music, Drama
 Biological Sciences, e.g. Chemistry, Biology
 Arts subjects, e.g. English Literature, History
 Social Sciences, e.g. Sociology, Psychology
 Geography and related subjects, e.g. Geography
 Economics and related subjects, e.g. Politics
 Mathematics and related subjects, e.g. Statistics
 Physical Sciences, e.g. Physics, Chemistry
 Engineering, e.g. civil or mechanical engineering

- a. Do you think you would be more successful at these subjects than the others?
 - b. Why?
 - c. Do you think your male classmates would choose the same subjects?
 - d. What about your female classmates? What subjects would they choose?
- Prompt: Who would choose language learning?
- a. Why?
 - e. Do you think it is appropriate for males/females?
 - f. Do you have male English teachers?
 - g. If language learning is a feminine domain, how do you think they managed to learn English?

6. Do you think your teachers (also) think that language learning is a feminine domain? Prompt: Do you think your teachers favour females more than males in foreign language classes?
7. Do you think your teachers communicate with female and male students differently because of this belief?
 - a. How?
 - b. Why?
 - c. Any examples?

Prompt: Do your language teachers have different performance expectations with males and females?

 - a. How different are the expectations?

b. Males and Self-Efficacy Beliefs

8. Could you please name the most successful three students in your class?
 - a. Why do you think these students are successful?
 - b. What do they have in common?
9. What do you think unsuccessful students have in common?

10. If you compare your male classmates with your female classmates, who do you think are more confident in language classes?
- a. Why do you think so?
 - b. Any examples?
11. What about you? Do you believe that you can learn English easily?
- Prompt: Are you confident in learning English?
- a. Why? Why not?

c. Males and Foreign Language Anxiety

12. If you compare your male friends with your female classmates, who do you think are more anxious during the classes?
- a. Why?
 - b. Any examples?
13. What about you? Do you feel always anxious in language classes?
- a. Why?
 - b. How do you cope with your anxiety?
 - c. Do your teachers help with that?
 - d. How?
14. Do you think foreign language anxiety affects your language learning performance?
- a. How?
-

ENGLISH LANGUAGE TEST



LISTENING PART

TIME

- ✓ Approximately 20 minutes.

INSTRUCTIONS TO STUDENTS

- ✓ Do not open this question paper until you are told to do so.
- ✓ Listen to the instructions for each part of the paper carefully.
- ✓ Answer all the questions.

INFORMATION FOR STUDENTS

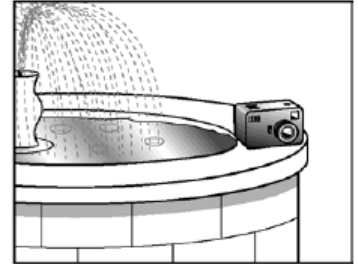
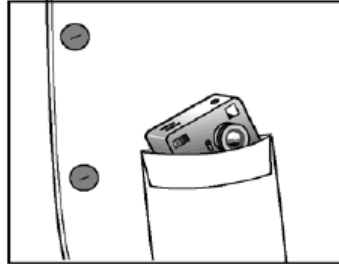
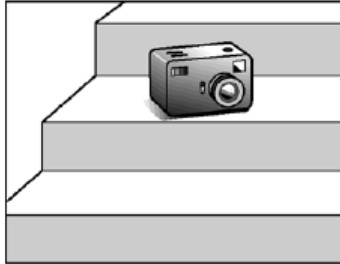
- ! There are four parts to the test.
- ! For each part of the test there will be time for you to look through the questions and time for you to check your answers.

Part 1

Questions 1 – 7

There are seven questions in this part.
For each question there are three pictures and a short recording.
Choose the correct picture and put a tick (✓) in the box below it.

Example: Where did the man leave his camera?

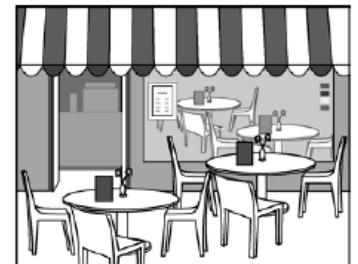
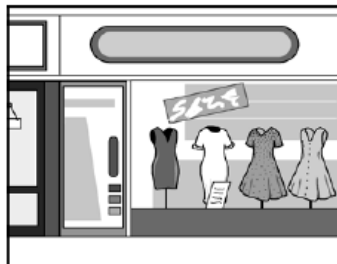


A ☒

B ☐

C ☐

1 Where will the women meet tomorrow?



A ☐

B ☐

C ☐

2 When will the man go to see the dentist?

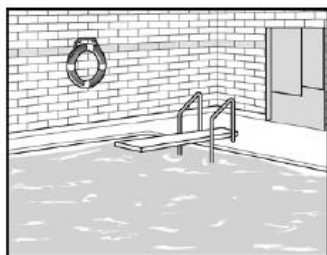


A ☐

B ☐

C ☐

3 Where are they at the moment?



A ☐



B ☐



C ☐

4 Where did the man stay on holiday?



A ☐



B ☐



C ☐

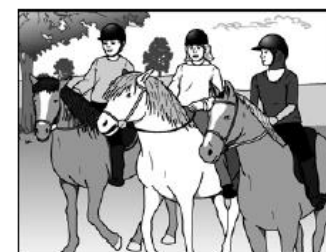
5 Who is the man going to work with?



A ☐



B ☐



C ☐

6 Where is the boy at the moment?



A ☐

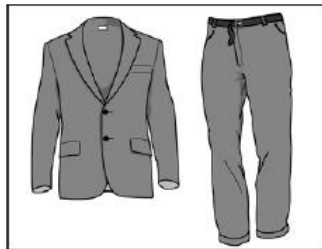


B ☐



C ☐

7 Which goods are reduced in price in the store now?



Part 2

Questions 8 – 13

You will hear an interview with a writer called Peter Taylor.
For each question, put a tick (✓) in the correct box.

- | | | |
|---|---|--|
| 8 In Peter's first book, the story takes place in | A a country which he's recently been to.
B a country where he lived as a child.
C the country where he was born. | <input type="checkbox"/>
<input type="checkbox"/>
<input type="checkbox"/> |
| 9 When Peter first went to England, he visited his | A grandfather.
B uncle and aunt.
C cousins. | <input type="checkbox"/>
<input type="checkbox"/>
<input type="checkbox"/> |
| 10 Peter spends most of his year | A near the beach.
B in a city.
C at his farm. | <input type="checkbox"/>
<input type="checkbox"/>
<input type="checkbox"/> |
| 11 What problem did Peter have in the desert? | A His vehicle broke down.
B He didn't have enough water.
C He was frightened by an animal. | <input type="checkbox"/>
<input type="checkbox"/>
<input type="checkbox"/> |
| 12 In his spare time, Peter usually | A goes to the cinema.
B gets together with friends.
C does photography. | <input type="checkbox"/>
<input type="checkbox"/>
<input type="checkbox"/> |
| 13 What does Peter want to do in the future? | A publish another novel
B write a history book
C spend more time travelling | <input type="checkbox"/>
<input type="checkbox"/>
<input type="checkbox"/> |

Please turn over for Part 3 ►

Part 3

Questions 14-19

You will hear a radio announcement about weekend activities in Fishport.
For each question, fill in the missing information in the numbered space.

WEEKEND ACTIVITIES IN FISHPORT

YOUR CHOICE OF ACTIVITIES

- BIRD ISLAND WALK – 10.00 a.m.

Don't forget to take your **(14)** with you

- GUITAR DAY

Will be held in the **(15)** Centre

- PLAZA CINEMA – 2.30 p.m.

A programme of **(16)** films for all the family

- CYCLE RACE

This year's route is through the **(17)**

- GREEN STREET THEATRE – 3.00 p.m.

'The Long **(18)** ' – a play for children

- CAMFORD CASTLE – open all day

Display of **(19)** used in medicine

Please turn over for Part 4 ►

Part 4

Questions 20-25

Look at the six sentences for this part.

You will hear a boy, Ian, and a girl, Sally, talking about cooking.

Decide if each sentence is correct or incorrect.

If it is correct, put a tick (✓) in the box under **A** for **YES**. If it is not correct, put a tick (✓) in the box under **B** for **NO**.

	A YES	B NO
20 Sally knows that Ian is an excellent cook.	<input type="checkbox"/>	<input type="checkbox"/>
21 Sally is happy to eat less meat than she used to.	<input type="checkbox"/>	<input type="checkbox"/>
22 Ian learned about cooking by watching other people.	<input type="checkbox"/>	<input type="checkbox"/>
23 Ian and Sally agree that schools should offer more cooking classes.	<input type="checkbox"/>	<input type="checkbox"/>
24 Sally is willing to pay more for dishes that are already prepared.	<input type="checkbox"/>	<input type="checkbox"/>
25 Ian suggests that simple recipes are best.	<input type="checkbox"/>	<input type="checkbox"/>